UNCOVERING FUNCTIONAL MECHANISMS IN CANCER THROUGH INTEGRATIVE GENOMICS

BANGARUSAMY DHINOTH KUMAR
M.Sc. (Biotech.), Madurai Kamaraj University, M.S (Mol.Bio.), NUS, Singapore

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOCHEMISTRY

NATIONAL UNIVERSITY OF SINGAPORE

February 2008

ACKNOWLEDGEMENTS

My first and sincere gratitude goes to my **"GURU" Dr. Lance David Miller** for his constant and unabated scientific guidance. His encouragement and enthusiasm has helped me to overcome several tough situations. Special thanks for his caring nature, patience and unwavering support through out the course of my graduate studies.

I am most grateful and indebted to my supervisor, **Prof. Edison Liu** for accepting me as his student and providing guidance and advices amidst his busy schedule. I have gained a lot from his sharp scientific intellect.

I am thankful to **Dr.Prasanna Kolatkar** for his support and supervision. With out him I would not have got a chance to pursue my doctorate degree at GIS.

I would like to express my gratitude to the past and present members of the lab for their constant help in one way or other during my stay at GIS. My heartfelt thanks to **Dr. Nallasivam Palanisamy**, **Dr. Krishnamurthy**, **Dr. Sabry** and **Vega** for their scientific and technical help.

My sincere thanks to all my friends especially Venthan, Srini, Siva, Saravanan, Srini (TLL) for their support and encouragement when I needed the most.

Special thanks to my beloved wife for her patience, love and perseverance during the stressful times (when I was being difficult), with out her support I would not have come this far. My family members (Amma, Dad, my brothers Vinoth and Asok and my in-laws) deserve heartfelt thanks for their trust and constant support through out these years.

A huge thanks to my precious son, Sai Sudhish who has given a new meaning to my life.

Dhinoth Kumar B.
February 2008

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

bp	base pair
cDNA	complementary DNA
DTT	dithiothreitol
EDTA	ethylene diamine tetra acetic acid
HCl	hydrochloric acid
Kb	kilobase
mg	milligram
ml	millilitre
min	minute
PBS	phosphate buffered saline
PCR	polymerase chain reaction
µl	microlitre
µM	micromolar
CO_2	carbon dioxide
PWM	position weight matrix

SUMMARY

The genome has been called the blueprint of life for it encodes a complete set of genetic instructions that specify the precise design and timing of functional molecules (such as RNAs and proteins) responsible for carrying out all cellular processes. In recent years, the human genome, comprised of approximately 3 billion nucleotide base pairs, has been decoded and determined to encode approximately 30,000 genes. This detailed genetic information has enabled the creation of advanced genomic technologies such as DNA microarrays that interrogate the structural and transcriptional dynamics of the genome on a comprehensive scale. However, the computational analysis of the output of genome-scale investigations has not been readily intuitive or subject to standardization. In this work, we have focused on the applications of genomic technologies towards the elucidation of cancer-related biomechanisms. From the development and coupling of analytical methodology and experimental design, to the prediction of genomic alterations from transcriptional measurements, this thesis describes a body of work aimed at extracting new fundamental insights into the pathobiology of cancer. Our genome-centric strategies and resulting cancer discoveries are presented.

Chapter 1: Understanding Cancer from a Genomic Perspective

1.0 Basis of cancer

Cancer is a disease of the genome and the genesis and malignant progression of cancer is initiated by genetic insults to the DNA such as sequence mutations, structural chromosomal alterations and epigenetic modifications. In early tumorigenesis, DNA damage arising from the harmful effects of genotoxic agents such as ionizing radiation, intercalating agents and free radical oxidants, can negatively affect the coding and/or regulatory regions of certain genes, giving rise to mutant or dysregulated proteins with a gain of "oncogenic" function (i.e., oncogenes) or with loss of tumor inhibitory activity (i.e., tumor suppressor genes). Alterations in these key genes spark a process of tumor progression marked by cumulative mutational events that invoke the inappropriate activation (or silencing) of a number of growth-regulating signaling pathways. Ultimately, these alterations converge on the selection of cells with certain malignant attributes such as the ability to replicate limitlessly, an affinity towards positive growth signals, disregard for growth suppressive signals, evasion of apoptosis, sustained angiogenesis and an ability to invade surrounding tissues and metastasize to distant organ sites [1].

At the protein level, oncogenes and tumor suppressors act in a variety of cellular contexts and subcompartments to modulate growth regulatory mechanisms in cancer. The epidermal growth factor receptors, EGFR/erbB and HER2/neu (ERBB2), localized at the plasma membrane, are classic examples of cell surface receptors whose activation by ligands or genomic amplification and subsequent over expression results in proliferative, migratory and survival advantages that lead to tumor metastasis [2-4].

The alteration of cytoplasmic components that receive and transduce signals originating from growth factor receptors can also contribute to cancer formation and progression. In the SOS-Ras-Raf-MAPK cascade, for example, mutations in the Ras oncogene, a GTPase whose physiologic role is to transduce extracellular signals in the MAPK (and PI3K) signaling pathways, can cause the constitutive activation of Ras resulting in the sustained transmission of a growth stimulatory signal that drives cellular proliferation. This cascade is connected to several downstream effector pathways and hence plays a central role in effecting tumorigenesis [5]. Activating Ras mutations are found in about 50% of colon carcinomas, 30-50% of lung adenocarcinomas, and more generally, in ~25% of all human cancers [6].

Alterations of growth negating signals that instruct cells to stop proliferating also occur, and these alterations affect tumor suppressor genes whose impact is often at the level of gene expression control. The Retinoblastoma gene (RB), discovered as the primary cause of the rare childhood eye tumor, retinoblastoma, is a central component in a pathway through which most anti-growth signals are transduced. The Rb pathway modulates the E2F transcription factors that in turn induce the expression of major cell cycle genes. Hyperphosphorylation of the Rb protein inhibits its sequestering activity over the E2Fs, thereby allowing E2F activity to drive uncontrolled cell proliferation [7]. The pathway can also be disrupted by a defective transforming growth factor beta (TGFβ) which transmits anti-proliferative signals through Rb. Malfunctions in other downstream elements can also converge on the RB pathway, further suppressing its growth inhibitory signals [8].

The discovery of the genes and pathways responsible for aberrant growth signals leading to cancer formation and progression is not only important academically, for a better understanding of cancer biology, but also is vital to the clinical goals of improved disease subtyping, patient prognosis and targeted therapeutics [9]. Clues to where these genes lie have historically been gathered through observation of the gross alterations in the cancer genome[10-12]). These alterations take several forms, and are frequently described as amplifications (copy number gains), deletions (copy number losses), and balanced rearrangements (such as translocations, inversions and transpositions that involve the exchange of chromosomal material between two chromosomes or intrachromosomal regions) [13].

Recurrent chromosomal aberrations have been reported in virtually all cancer types, often with clinical correlations. For example, in colorectal cancer it has been shown that patients with losses in chromosomal arms 1p, 4q, 8p, 14q, or 18q, or amplification of chromosomal arm 20q have poorer survival outcomes than those lacking these aberrations [14]. It has been estimated that more than 100 cancer-related genes have been identified through the discovery of nearly 600 recurrent balanced chromosomal alterations. Balanced chromosome rearrangements either deregulate genes at their breakpoints or result in a fusion gene by a process of recombination. In many instances, these breakpoints harbor either an oncogene that gets activated, or a TS gene that gets inactivated, as a consequence of the rearrangement. For example, in Burkitt's lymphoma three translocations, t(8;14), t(8;22), and t(2;8), juxtapose the MYC gene at 8q24 to the constitutively active immunoglobulin genes (i.e., IGH at 14q32, IGL at 22q11 and IGK at 2p12) resulting in Myc activation [15].

Epigenetic modifications, such as the hypermethylation-mediated transcriptional silencing of tumor suppressor genes, is another common etiology of cancer [16, 17]. Recent studies have shown the involvement of hypermethylation in major pathways leading to cancer such as DNA repair (BRCA1, MGMT), cell cycle regulation (p14, p15, p16), apoptosis (DAPK, APAF-1), carcinogen metabolism (GSTP1), hormonal response (RARβ2) and cell adherence (CDH1, CDH3) [18-20]. Thus, cancer is a complex and heterogeneous disease, driven by a multitude of inappropriate signals with genomic alterations at their origin [21].

1.1 Historical views of the cancer genome

As early as 1890, before the concept of the gene, clues to the connection between the genome (chromosomes) and cancer were observed when Hansemann postulated that asymmetric mitoses observed in some tumor cells were at the root of cancer formation [22]. In 1928, with the rediscovery of Mendel's work, genetic mutations were proposed as the origin of cancer, as an alternative to the infection theory of cancer [Bauer]. However, it would not be until 1969 that a technique would be developed to allow the recognition and tracing of individual chromosomes. Called chromosome banding, this approach was a major breakthrough in the field of cancer genomics as it enabled the study of chromosomal copy number and rearrangements. In this technique, chromosomes were chemically labeled with Giemsa stain (specific for DNA phosphate bonds) following a trypsin digestion resulting in a series of dark and light bands representing the AT-rich heterochromatic regions and the GC-rich euchromatic regions, respectively. These bands enabled researchers to visualize and thus order the chromosomes into pairs - a procedure termed "karyotyping". The utility of this technique was definitively demonstrated when it was successfully used to discover the Philadelphia chromosome - a

reciprocal translocation designated as t(9;22)(q34;q11) that is the hallmark of chronic myelogenous leukemia [23]. Other cytogenetic tools followed, including fluorescence in-situ hybridization (FISH), Spectral Karyotyping (SKY) and comparative genomic hybridization (CGH), that helped cancer researchers locate regions of interest in the cancer genome. These techniques allow the reliable identification and characterization of complex chromosomal rearrangements and quantitative copy number changes.

1.2 Modern methods of characterizing the cancer genome

1.2.1 Fluorescence in-situ hybridization (FISH)

Fluorescence in-situ hybridization (FISH) was the first of the so-called "modern" cytogenetic techniques developed as an adjunct to the classical cytogenetic analysis [24]. In FISH, DNA probes of known origin are fluorescently labeled and hybridized to a metaphase chromosome spread or interphase nuclei. The probes anneal to the homologous sequences within the chromosome and are detected via fluorescence microscopy. Thus, FISH allows the quantitative determination of copy number status (i.e., amplification or deletion). For a specific chromosomal region, it provides visual information on which chromosome(s) received the amplified region (know as an amplicon). This technique is effectively used to study whole chromosomes or specific regions such as centromeres, telomeres, or specific genes, as well as juxtaposed chromosomal aberrations.

1.2.2 Spectral karyotyping

Spectral karyotyping or SKY is a multi-chromosomal painting assay that was developed to help visualize and distinguish chromosomes from one another. Like FISH, SKY is also a

fluorescence-based technique wherein fluorescently-labeled, chromosome-specific, composite probe pools are prepared and hybridized to metaphase or interphase cell spreads. Importantly, each probe pool incorporates a different fluorescent label, thus allowing the colorimetric discernment of individual chromosomes. This technique has become an indispensable tool for both basic research and diagnostic applications[25], as it provides a rapid means of identifying the numerical and structural chromosomal aberrations that characterize the cancer genome.

1.2.3 Comparative genomic hybridization

Comparative Genomic Hybridization (CGH) is a FISH-related technique developed to facilitate the ability to scan the entire cancer genome in an unbiased fashion for changes in DNA copy number[26]. In this technique, fluorescently labeled tumor DNA (e.g., fluorescein (FITC)) and normal DNA (e.g., rhodamine or Texas Red) are hybridized to normal human metaphase preparations. With the aid of epifluorescence microscopy and quantitative image analysis, regional differences in the fluorescence ratio indicating the gain or loss of tumor DNA can be detected and used for identifying abnormal regions in the genome. This method can only identify the copy number changes in the genome and hence is not useful for discovering balanced chromosomal changes like reciprocal translocations and inversions. While this technique has been effectively applied to cancer resulting in the discovery of genomic changes of clinical importance, the resolution of conventional CGH is poor, approximately 10Mbp, limiting its view to only macro-scale genomic alterations[27].

1.2.4 Microarray-based comparative genomic hybridization

DNA microarrays are ordered collections of tens to hundreds of thousands of DNA (probe) sequences spotted onto a solid substrate such as chemically modified glass. Typically, these DNA probes are designed to be complementary to genes or genomic sequences, and upon hybridization with fluorescently labeled "target" nucleic acid, will provide a semi-quantitative readout of specific nucleic acid levels. In the case of microarray comparative genomic hybridization (array CGH), the microarray probes can be designed to recognize individual genes (in the form of cDNA PCR products or synthetic oligonucleotides) [28-31] or long tracts of genomic sequence (such as bacterial artificial chromosome (BAC) clones) [32-34]. To this array, differentially fluorescently labeled genomic DNA from "normal" and cancerous tissue can be hybridized, generating a fluorescence "signal" ratio that reflects relative changes in genomic copy number between the cancer and normal tissue [35, 36] .

Compared to traditional CGH which identifies genomic alterations on a "macro" chromosomal scale, array CGH has the advantage of substantially higher resolution, as defined gene/genomic sequences are used to measure copy number status. Recently, state-of-the-art high-density arrays comprised of $>1x10^6$ tiled oligonucleotide probes have been reported to achieve genomic resolution of <20 kilo bases [37]. This high level of resolution allows the discovery of "micro" amplicons and deletions that may alter only a single gene [38].

In addition to microarrays for studying the genome, microarrays for studying the transcriptome have also become a valuable molecular tool. The *expression array* measures gene activity or more specifically, mRNA levels on a genome-wide scale. Thus, the DNA probes of an

expression microarray are design to hybridized to the transcribed regions of genes, and fluorescently labeled cDNA (or cRNA) derived from cellular mRNA, is the target for an expression array hybridization [31].

1.3 Towards cancer gene discovery

Cytogenetic chromosomal mapping strategies such as FISH, SKY and CGH have traditionally been utilized as a starting point for uncovering oncogenes and tumor suppressor genes at genomic sites of recurrent amplification and deletion, as these regions are thought to harbor genes with causal roles in tumorigenesis. This approach has its limitations, however, as many candidate genes are often identified, and serial testing for the oncogenic properties of candidate genes is both costly and time consuming. Thus, new approaches are necessary for narrowing down the candidate gene lists, and recent studies suggest that the integration of multiple forms of information such as genomic, transcriptomic, biological, and even clinical data may provide a more successful strategy [39, 40].

1.4 Integrative genomic analysis

Recent results from several laboratories suggest that combining microarray gene expression measurements with genome copy number data can enhance the cancer gene discovery process. While recurrent CNAs are believed to alter the expression levels of key cancer-promoting genes, not all genes residing at a given CNA are necessarily dysregulated at the transcriptional level. By integrating CNA data from array CGH analysis with expression data from expression array analysis, one can study the correlations between copy number change and the transcriptional modulation of genes leading to the generation of candidate gene lists that can

be narrowed down by filtering for only those that show a statistically significant or "best" correlation [41]. The intersection with clinical correlations can provide even further filtering for narrowing down candidates. Studies that rely on this principle of genomic and transcriptomic integration are generating new insights into mechanisms of human diseases including cancer [39].

In a recent study by Chin et al, the combination of gene expression with copy number data was employed to improve the classification of patients according to clinical outcome [41]. The study demonstrated that the genes deregulated by the recurrent genome CNAs serve as efficient biomarkers to determine the treatment regime of patients. Furthermore, re-classification of patients with basal-like tumors using this approach showed better outcome than the previous classifications that were based on expression patterns alone. This study also re-iterated the nexus between certain high-level amplifications and reduced survival duration. Specifically, the study identified 66 genes whose expression levels correlated with copy number and patient survival suggesting their utility as therapeutic targets - nine genes (*FGFR1*, *IKBKB*, *ERBB2*, *PROCC*, *ADAM9*, *FNTA*, *ACACA*, *PNMT*, and *NR1D1)* are currently considered druggable. Furthermore, an association between low-level CNAs and upregulation of RNA and protein metabolic genes that contribute towards cancer progression was established using this comprehensive approach. Thus, integrative analysis of genomic alterations and transcription profiles has the potential to reveal important genes and clinical associations with prognostic, diagnostic and therapeutic implications in cancer.

In the following chapters of this thesis, I will describe how we have developed and applied novel integrative concepts for mining genomic and transcriptomic data to uncover cancer related genes and pathways using an experimental model of nasopharyngeal carcinogenesis.

Chapter 2: Integrating Genomic Analysis and Experimental Methodology to Uncover Tumor Suppressor Genes in Nasopharyngeal Carcinoma

2.0 Nasopharyngeal carcinoma (NPC) and evidence for tumor suppressor gene activity

Nasopharyngeal carcinoma (NPC) is a rare form of cancer that originates from the epithelium of the nasopharynx and is generally associated with poor patient prognosis [42]. NPC has the highest occurrence in Southern China, and the prevalence of this disease particularly among the Cantonese Chinese suggests that there may be a genetic predisposition to NPC[43]. While Epstein-Barr virus (EBV), diet and environmental factors are thought to contribute to NPC progression [44, 45], the molecular events that lead to NPC are very poorly understood. Several cytogenetic studies involving NPC specimens have reported frequent chromosomal aberrations and loss of heterozygosity at 3p, 9p, 11q, and 14q [46-49]. It has been hypothesized that the inactivation of tumor suppressor genes (TSGs) present in these chromosomal regions may contribute directly to NPC progression. TSG activity at 11q13 has been observed in several other cancers types. Notably, the *MULTIPLE ENDOCRINE NEOPLASIA* type I (*MEN1*) gene has been mapped and cloned from this region [13, 50]. However, *MEN1* expression does not affect growth of NPC cell lines[51]. TSG activity at 11q22-23 has also been observed in several cancers such as melanoma[52], breast[53], ovarian[54], lung[55], cervical[56], bladder[57], colorectal[58], and prostate cancer[59]. A novel TSG, *TUMOR SUPPRESSOR IN LUNG CANCER 1* (*TSLC1*), located in this region has been identified in non-small cell lung cancer by functional complementation[60]. However, a role for *TSLC1* in NPC progression has not been clearly defined [61]. Finally, molecular and cytogenetic analyses have also revealed extensive losses at 11q22-24 in NPC[62] but no strong candidate TSGs from this region have yet been

proposed. Together, these findings suggest that one or more TSGs located on chromosome 11q play a critical role in NPC development.

2.1 Microcell-mediated chromosome transfer and discovery of chromosome-specific TSGs.

Microcell-Mediated Chromosome Transfer (MMCT) (also known as somatic cell hybridization) is a chromosomal transfer technique developed for the identification of disease-bearing chromosomes [63, 64]. The MMCT procedure begins with a series of sub cellular manipulations designed to generate single-chromosome bearing microcells derived from a donor cell line. The chromosome-bearing microcell can then be fused to a recipient cell line, which can be further selected for cell hybrids that contain the chromosomes of interest, or in some cases, a phenotype of interest [65]. In the context of cancer, where it may be hypothesized that a particular chromosome or chromosomal fragment harbors a tumor suppressor gene, a functional complementation approach can be taken. MMCT can be used to introduce “normal" chromosomal DNA into a malignant cell lacking the chromosomal DNA of interest followed by screening for reversion of the tumorigenic phenotype [65, 66].

In the current work, our collaborators previously used MMCT to introduce an intact chromosome 11 into a tumorigenic NPC cell line referred to as *HONE1*, generating a number of “hybrid” cell lines that exhibited a suppression of tumorigenicity when passaged in nude mice[51]. Presumably, the intact chromosome 11 was able to complement the tumor suppressive functions of the defective chromosome 11 of the host cell. Our collaborators then observed that these hybrid cells remained non-tumorigenic for approximately three months, at which point they

regained tumorigenic behavior. These tumorigenic cells, termed “segregants”, thus exhibited a delayed latency period in tumor formation compared with that of the parental HONE1 cells [67] (Figure F2.1). The delayed tumor appearance was presumably the result of loss or inactivation of tumor suppressive wild type alleles present on the normal donor chromosome 11.

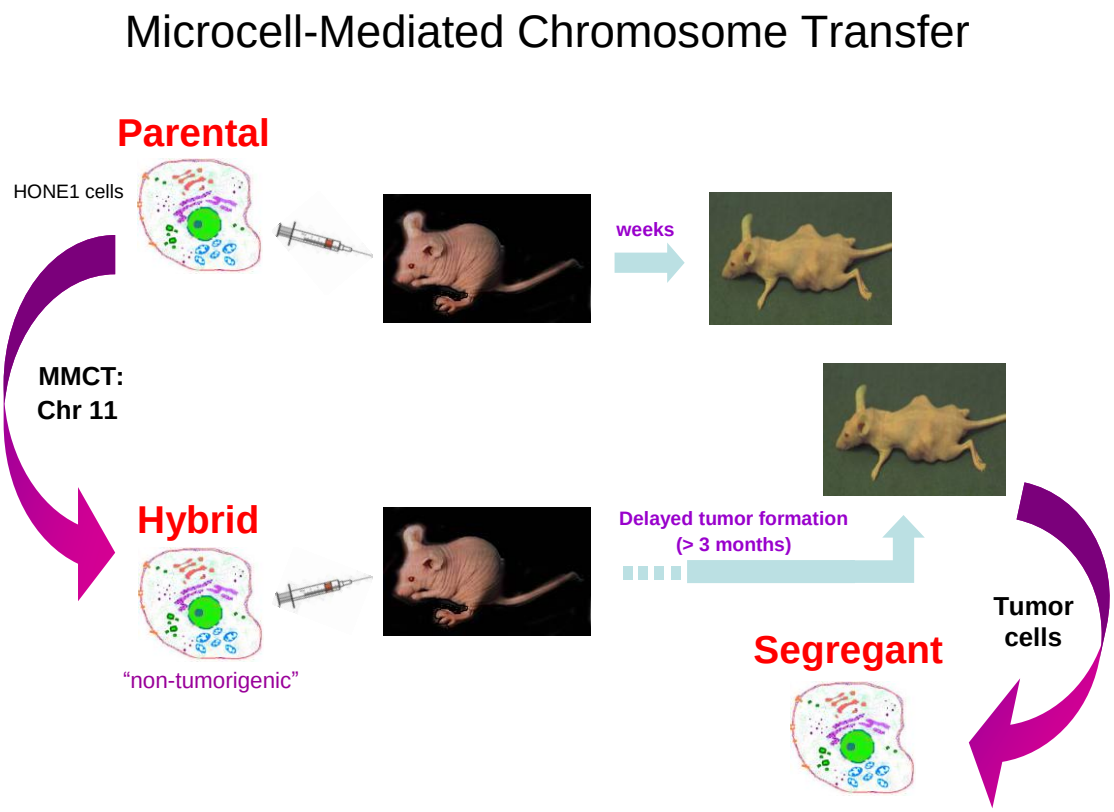


Figure F2.1: Experimental strategy for identifying TSG-bearing chromosomal regions.

The diagram explains the strategy used for deriving hybrid and segregant cell lines from the parental HONE1 cells.

2.2 MMCT with a genomic twist

Several tumor segregant cell lines were isolated from the mice and cultured for genetic analysis of chromosome 11. Using detailed comparative BAC FISH and microsatellite marker analyses of non-tumorigenic hybrid and tumor segregant cell lines, critical genomic regions lost in nasopharyngeal carcinoma (NPC) were identified including a 1.8 Mb region at chromosome band 11q13 and three other regions of 0.36 Mb, 0.44 Mb, and 0.3 Mb at 11q22-23 [51, 67].

While these "critical regions" were consistently deleted from the segregant cell lines, the tumor suppressor gene(s) driving the observed phenotypes could not be readily deduced as greater than 100 known genes mapped to these critical regions. We sought to address this problem with expression microarray analysis. We hypothesized that while the expression of all genes localized to the critical regions would be suppressed (or absent) in the parental line and segregants, the key tumor suppressor gene(s) would be transcriptionally active in the non-tumorigenic hybrid lines. Thus, we posited that the "signature" of a tumor suppressor gene should comprise of the consistent *inactivation* of expression in the both the parental HONE1 cells and the segregant cell lines, and consistent *activation* of expression in the hybrid cell lines.

2.3 RESULTS

2.3.1 Microarray analysis in HONE1 cells, hybrids and tumor segregants

Using a 19K oligonucleotide expression microarray containing probes to most human genes and all known genes on chromosome 11q, gene expression profiles of parental, hybrid and segregant cell lines were analyzed by competitive hybridization. Specifically, mRNA of the tumorigenic parental HONE1 line was hybridized against each of four independent (non-tumorigenic) hybrid lines (HK11.8, HK11.12, HK11.13, and HK11.19), and each of the four hybrid lines was hybridized against its corresponding tumor segregant (HK11.8-3TS, HK11.12-2TS, HK11.13-1TS, and HK11.19-4TS). A diagram of the experimental strategy is shown in Figure F2.2A & F2.2B. As each hybridization experiment was performed in duplicate (i.e., in dye-swap fashion), sixteen microarray experiments were performed in total (Appendix I 1.0 & 1.1).

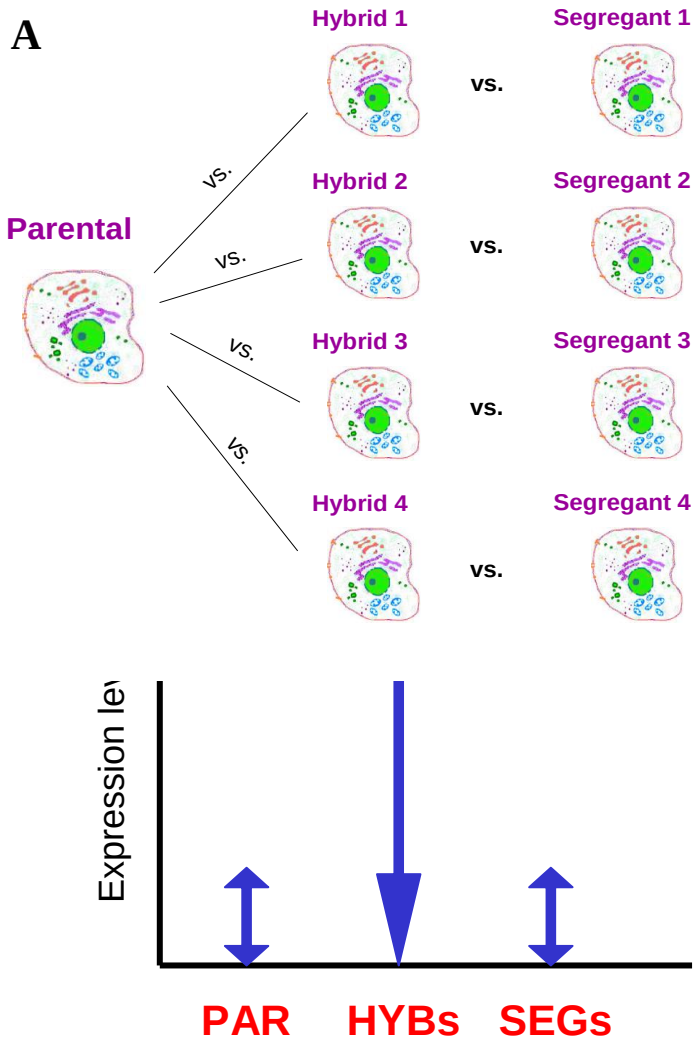


Figure F2.2: Microarray comparisons and the theoretical tumor suppressor signature.

A) The diagram shows the microarray strategy adapted for the experiment. The HONE1 parental cell line was hybridized against each of four hybrid lines (biological replicates), and in turn, each of the hybrid lines was hybridized against its corresponding segregant line.
B) The indicative expression signature of a tumor suppressor gene: low expression in the parental cell line (PAR), activated expression in the hybrids (HYBs), and low (or absent) expression in the segregants (SEGs).

The data was then analyzed to identify genes up-regulated in the hybrids (after chromosome transfer) and down-regulated in the HONE1 line and tumor segregants (after loss of chromosome 11 material) (Figure F2.2B). Genes fitting this prescribed behavior were selected based on the correlation between their expression pattern and the theoretical tumor suppressor "signature" pattern (binarized as 1=up-regulated and -1=down-regulated). Genes with a Pearson correlation score >0.9 were selected as top tumor suppressor candidates, regardless of their genomic location. In total, 24 genes met this selection criterion, one of which mapped to a critical region of 11q at band q22.3 (Table T2.1). This gene, named *THY1*, was found differentially expressed under our selection criteria in three out of the four hybrid-segregant pairs. The gene expression ratios of each duplicate experiment and the mean ratio of the duplicate experiments for the hybrid-parental and hybrid-segregant comparisons are summarized in Table 2. A comparison of the expression profile of HONE1 cells with the four hybrids, showed a nearly 2-fold increased expression of *THY1* in all the hybrids except HK11.13 (Table T2.2). Reciprocally, the mRNA levels of *THY1* were significantly decreased in the tumor segregants, again with the exception of the HK11.13-11.13-1TS comparison (Table T2.2). Thus, *THY1* emerged as our top candidate tumor suppressor gene in our model of NPC.

GENE SYMBOL & NAME	CYTOBAND
SAA2 serum amyloid A1	11p15.1
MAP1B microtubule-associated protein 1B	5q13
MAP1B microtubule-associated protein 1B	5q13
INSIG1 insulin induced gene 1	7q36
SQLE squalene epoxidase	8q24.1
FDFT1 farnesyl-diphosphate farnesyltransferase 1	8p23.1-p22
ARTN artemin	1p33-p32
HSD17B7 hydroxysteroid (17-beta) dehydrogenase 7	1q23
THY1 Thy-1 cell surface antigen	11q22.3
SAA1 serum amyloid A2	11p15.1
ACAS2 acetyl-Coenzyme A synthetase 2 (ADP forming)	20q11.23
NFKBIA nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	14q13
FLJ36031 hypothetical protein FLJ36031	7q22.2
TRAF1 TNF receptor-associated factor 1	9q33-q34
STARD4 START domain containing 4, sterol regulated	5q22.2
GCH1 GTP cyclohydrolase 1 (dopa-responsive dystonia)	14q22.1-q22.2
CNTNAP1 contactin associated protein 1	17q21
Homo sapiens mRNA; cDNA DKFZp686B15184 (from clone DKFZp686B15184	--
CDKN2D cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4)	19p13
CTSS cathepsin S	1q21
CYP51A1 cytochrome P450, family 51, subfamily A, polypeptide 1	7q21.2-q21.3
C14orf1 chromosome 14 open reading frame 1	14q24.3
IL3RA interleukin 3 receptor, alpha (low affinity)	Xp22.3
MCFP mitochondrial carrier family protein	7q21.13

Table T2.1: Top TSG candidates.

Shown are the genes most highly correlated with the theoretical TSG expression signature, THY-1 cell surface antigen mapping to the 11q22.2 critical region is shown in red.

Cell lines	Expression ratios* Cy5/Cy3	Cy3/Cy5	Average ratio
11.8/HONE1	1.8702	2.7533	2.3
11.8-3TS/11.8	0.5686	0.5537	0.56
11.12/HONE1	2.3329	2.6434	2.5
11.12-2TS/11.12	0.4578	0.4509	0.45
11.13/HONE1	0.7835	1.4379	1.1
11.13-1TS/11.13	0.8892	1.058	0.97
11.19/HONE1	2.0286	1.9872	2.0
11.19-4TS/11.19	0.4820	0.5696	0.53

***The expression ratio for duplicate "dye-swap" arrays are shown: Cy5/Cy3 and Cy3/Cy5. For the Cy3/Cy5 dye-swap, the reciprocal expression ratio is shown.**

Table T2.2: Expression ratios of *THY1*.

The expression ratio is equivalent to fold-change; ratios >1 indicate higher expression in the hybrids, while ratios <1 indicate lower expression in the tumor segregants (TS). Note that a ratio of 0.5 is equivalent to a detected 2-fold decrease (in the tumor segregants).

2.3.2 Real-time PCR confirmation of THY1 expression patterns

To confirm the microarray findings, *THY1* expression levels were analyzed by real-time (RT)-PCR (Figure F2.3). RT-PCR is considered a much more quantitative measure of mRNA levels than microarray analysis (where seemingly small changes in expression levels by microarray can be, in reality, infinitely large as detected by RT-PCR). Quantification of *THY1* expression revealed a good correlation between the changes observed in the microarray experiments and the RT-PCR results. First, no expression of *THY1* was detected in the HONE1 cells, and *THY1* expression appeared to be absent in an additional three NPC cell lines, HK1, CNE1, and HNE1. Second, consistent with the microarray results, *THY1* was transcriptionally activated in the hybrids HK11.8, HK11.12, and HK11.19, but not in HK11.13. Finally, again consistent with the microarray observations, when comparing *THY1* expression between the hybrids HK11.8, HK11.12, and HK11.19 and their corresponding tumor segregants, RT-PCR revealed consistent loss of *THY1* expression in the tumor segregants (Appendix I 1.2).

2.3.3 Protein analysis of THY1 in NPC

Next, *THY1* protein levels were assessed by western blot (Figure F2.4). Similar to the findings from RT-PCR, HONE1, HK1, CNE1, and HNE1 did not express *THY1* protein, while protein was observed in the hybrids HK11.8, HK11.12 and HK11.19, as well as the chromosome 11 donor cells (MCH556.15). Interestingly, the HK11.13 hybrid, which did not express detectable levels of *THY1* mRNA, did express a low but detectable level of *THY1* protein. Additionally, *THY1* protein was not detectable in all four tumor segregants. Thus, the protein-level analysis of *THY1* expression is largely concordant with the real-time PCR results (Appendix I 1.6).

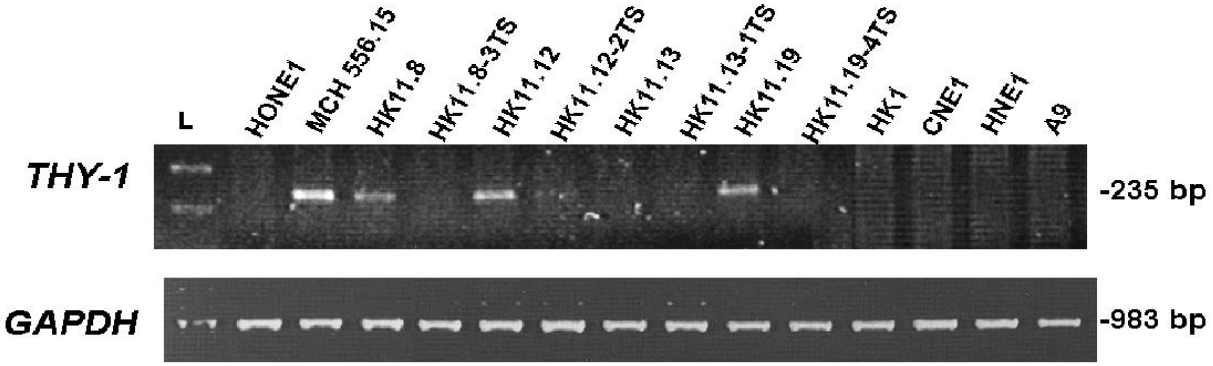


Figure F2.3: Real-time PCR analysis of *THY1*.

Real-time PCR analysis of *THY1* mRNA levels in the parental (HONE1), hybrid and segregant lines; 3 additional NPC lines (HK1, CNE1, HNE1); the mouse chromosome 11 donor cell line, MCH556.15, overexpressing *THY1* (positive control); and the mouse cell line A9 (negative control). The DNA ladder (L) indicates the sizes of the PCR bands, which are shown on the right (bp). *GAPDH* served as an internal control.

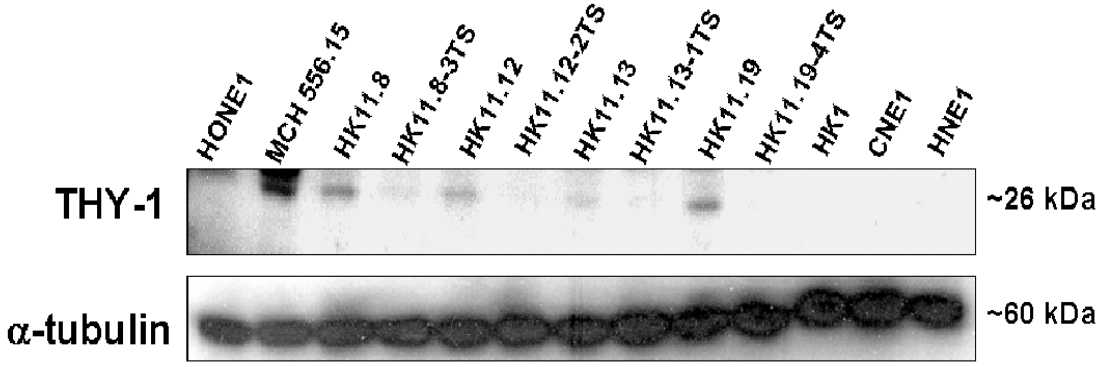


Figure F2.4. Western blot analysis of *THY1*.

The sizes of protein bands are indicated on the right. α-tubulin was used as an internal control.

2.3.4 THY1 promoter methylation in NPC cell lines

Notably, the 11q22.3 critical region harboring *THY1* was neither deleted in the parental HONE1 cells, nor several other NPC cell lines examined (data not shown). To determine the inactivation mechanism of *THY1* in NPC, we studied the methylation status of the *THY1* promoter in the four NPC cell lines by methylation-specific PCR (MSP) analysis. Figure F2.5A shows the detection of methylated sequences in HONE1, HK1, CNE1 and HNE1 cells, while a lesser extent of unmethylated signal was observed in the HONE1 and CNE1 cell lines (Appendix I 1.4). Subsequent re-expression of endogenous *THY1* in HONE1 cells was observed after treatment with the demethylating agent 5-aza-2'-deoxycytidine (Figure F2.5B). Together, these results suggest that methylation of at least one *THY1* allele is a common event in nasopharyngeal cancer.

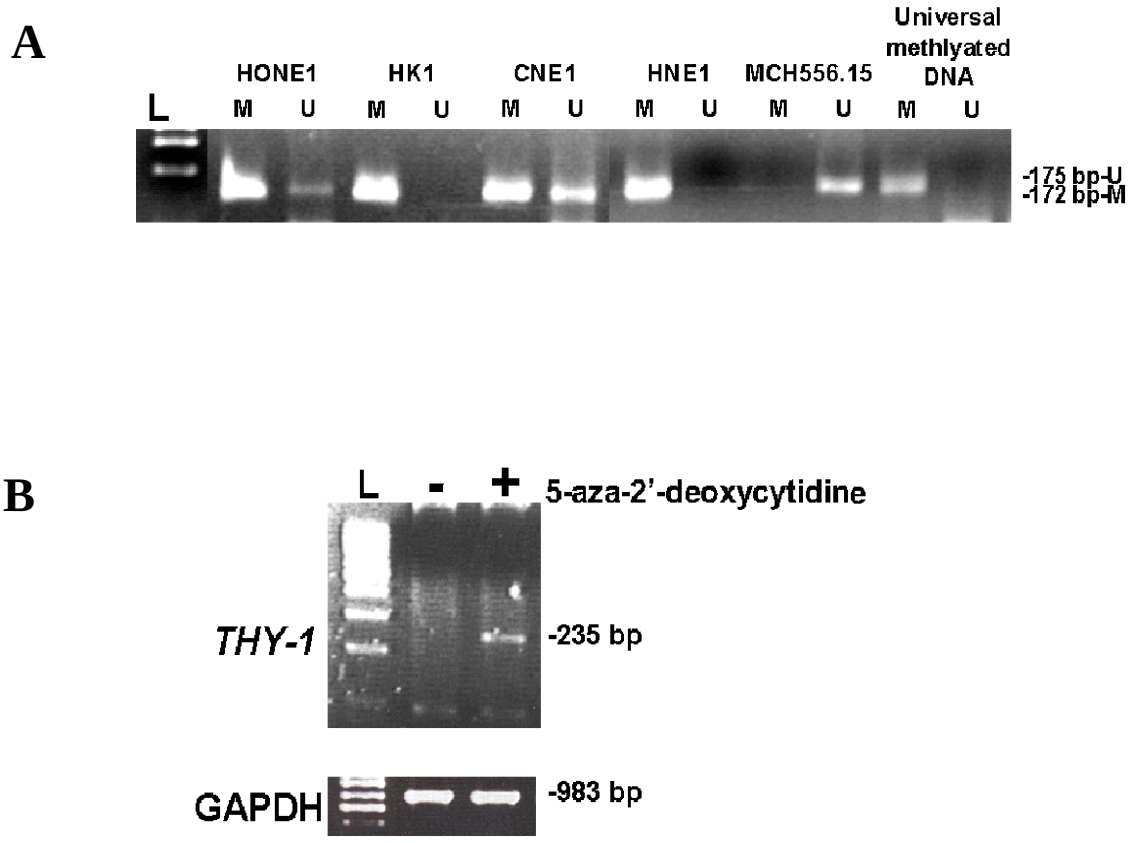


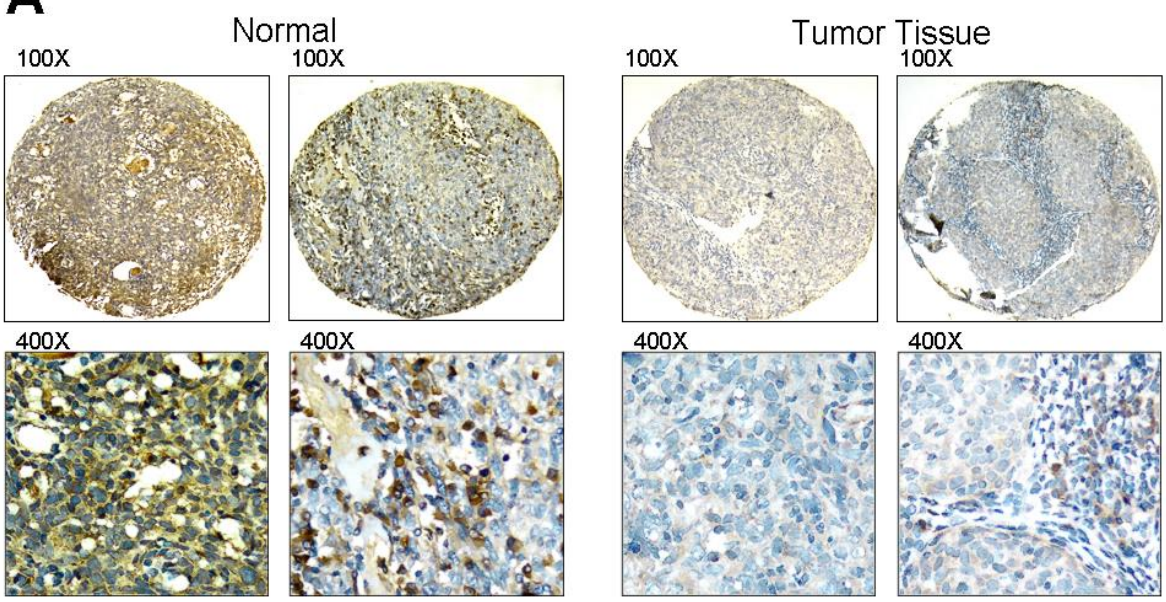
Figure F2.5. *THY1* methylation profiles.

A) MSP analysis of *THY1* promoter methylation in four NPC cell lines (HK1, CNE1, HNE1, and HONE1), MCH556.15, and the universal methylated DNA (positive control). Sizes of PCR products are shown on the right. A methylated allele was observable in all four NPC cell lines. **B)** Re-expression of endogenous *THY1* in HONE1 cells after treatment with 5 μM 5-aza-2'-deoxycytidine detected by RT-PCR.

2.3.5 Analysis of THY1 expression variation in human NPC samples

To investigate the natural variation of endogenous *THY1* expression in human nasopharyngeal carcinoma, we analyzed seventy clinical patient samples of NPC and nine samples of noncancerous nasopharyngeal mucosa for expression of *THY1* protein by immunohistochemistry on a tissue microarray (TMA) (Figure F2.6A) (Appendix I 1.7 & 1.8). We observed that the staining index of *THY1* expression in the noncancerous samples of nasopharyngeal mucosa showed an upper bound score of 6 or greater; therefore, we designated the staining index of 6-9 as the "normal" baseline expression of *THY1*. Accordingly, a staining index of 1-4 was considered as reduced expression, and a staining index of 0 was considered to reflect absence of expression. In the 70 NPC cases, 44% (31/70) showed down-regulated expression of *THY1*, while another 9% (6/70) were scored as absent for *THY1* expression. In the fraction of samples associated with lymph node metastasis at diagnosis, the frequency of down-regulated/absent *THY1* was 63% (17/27), significantly higher than that observed in primary NPC (33%) (14/43) ($P < 0.05$; Figure F2.6B). Thus, reduced expression of *THY1* was observed to be significantly associated with more aggressive NPC.

A



B

Expression of THY1 in primary and lymph node metastatic NPCs

	Normal expression	Down-regulated expression	Loss of expression	Total
Normal	9 (100%)	0 (0%)	0 (0%)	9
Primary NPC	26 (60%)	14 (33%)	3 (7%)	43
Metastatic NPC	7 (26%)	17 (63%)	3 (11%)	27
	$p < 0.05$	$p < 0.05$	not significant	

Figure F2.6. Immunohistochemical staining results for *THY1*.

Immunohistochemical staining of *THY1* in NPC TMA containing noncancerous nasopharyngeal mucosa, primary NPC, and lymph node-metastatic NPC.
A) Representative staining of *THY1* in normal nasopharyngeal mucosa, and tumor tissues showing reduced or loss of *THY1* expression.
B) The expression levels of *THY1* detected in normal mucosa, primary, and metastatic NPC are shown.

2.3.6 Functional analysis of THY1 growth suppressive properties

We next sought to functionally characterize *THY1* for growth suppressive activity using colony formation assays. HONE1 cells were transfected in replicate (4x) with a construct overexpressing *THY1* and neomycin resistance (pCR3.1-*THY1*) or the construct (with only neomycin resistance) alone (pCR3.1; as control). A significant decrease in the number of neomycin-resistant colonies was observed in the *THY1* overexpressing cells as compared to those transfected with the vector alone (Figure F2.7). Specifically, we observed a 90% decrease in colony formation in the *THY1* transfectants. To further test the inhibitory effect of *THY1* on colony formation, a recently developed tetracycline-repressible transgene system [68] was utilized. HONE1 cells expressing the tetracycline trans-activator tTA were established, and these cells (HONE1-2) were then stably transfected with the tetracycline-repressible construct containing the *THY1* sequence (pETE-BSD-*THY1*) or without the *THY1* sequence (pETE-BSD) (Appendix I 1.10). In the absence of doxycycline (a potent tetracycline derivative), *THY1* protein was exclusively expressed in HONE1-2 cells transfected with pETE-BSD-*THY1* (Figure F2.8A), and a significant reduction in the number of blasticidin-resistant colonies was observed (Figure F2.7 and 2.8A). However, with the addition of doxycycline, *THY1* expression was mitigated, and no significant change in colony number was observed between the pETE-BSD-*THY1* and pETE-BSD transfected HONE1-2 cells (Figure F2.8B). Together, these results provide direct evidence that *THY1* expression can inhibit growth of NPC cells, while *THY1* inhibition can augment NPC growth.

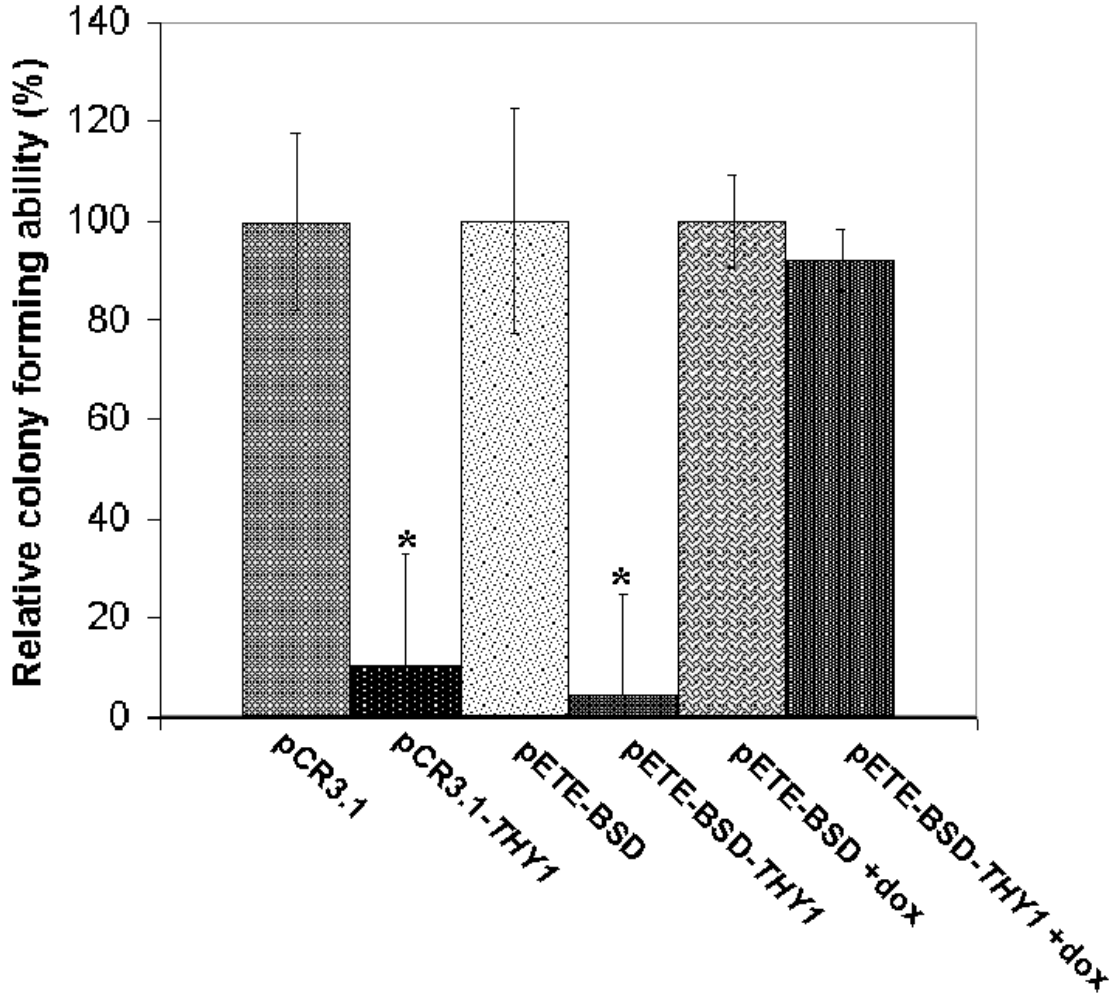


Figure F2.7. Colony formation assays with *THY1* transfectants.

HONE1 cells were transfected with pCR3.1-*THY1* or pCR3.1, and HONE1-2 cells were transfected with pETE-Bsd-*THY1* or pETE-Bsd in the presence or absence of doxycycline (dox). The colony forming ability was calculated by comparing the number of colonies in *THY1* transfectants to that of vector alone. Each treatment was performed in quadruplicate. * designates a significant difference from vector-alone ($p < 0.005$).

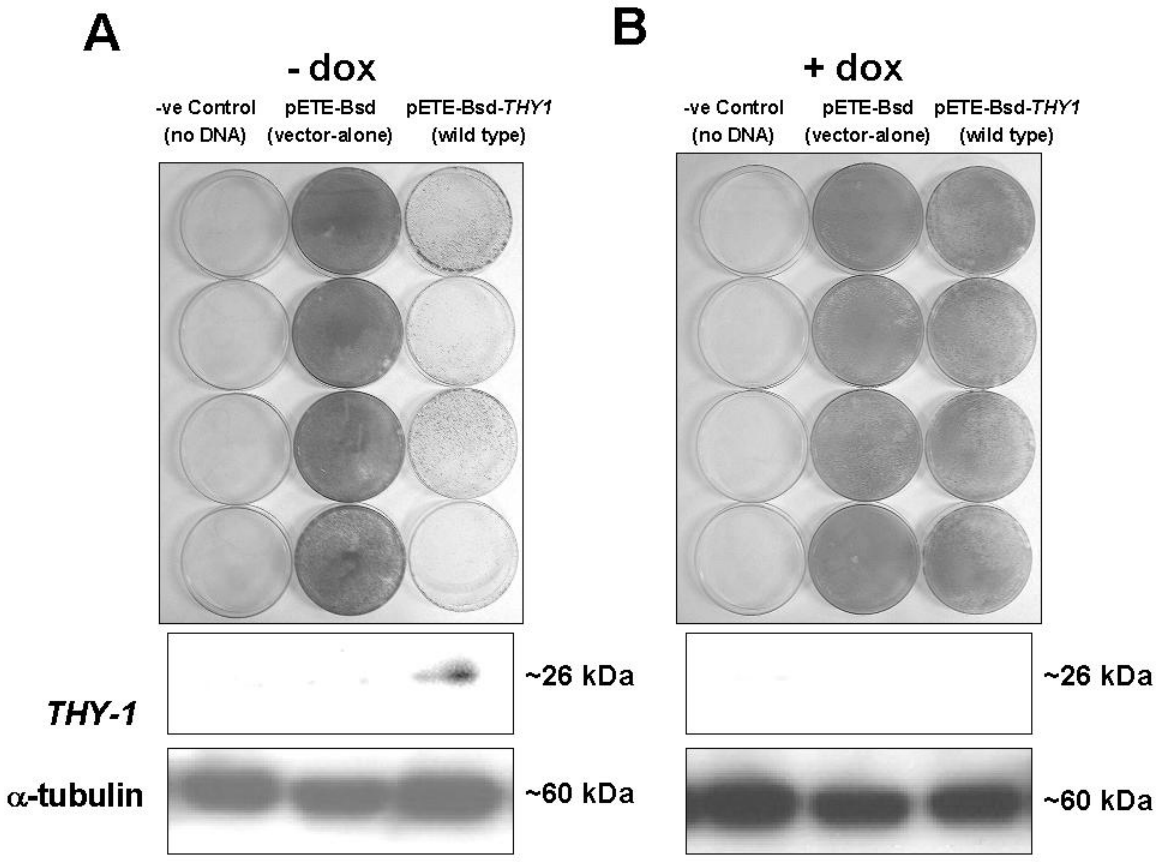


Figure F2.8: Growth effects of *THY1* expression in a Tet-repressible system.

Representative results of the colony formation assay and Western blot analysis of the *THY1*-transfected HONE1-2 cells and vector-alone transfectants. HONE1-2 cells were transfected in
A) the absence of dox, and
B) the presence of dox. Western blots (lower panels) and colony staining were performed 3 days post-transfection.

2.4 Discussion

Using a comprehensive oligonucleotide microarray platform, we analyzed the global gene expression patterns of a tumorigenic NPC cell line (HONE1), 4 independently-derived nontumorigenic hybrid lines (each containing an introduced normal copy of chromosome 11), and 4 corresponding tumorigenic segregant lines (derived from the hybrids) that share several critical regions of chromosome 11 deletion. We postulated that the signature of a tumor suppressor gene would be characterized by activated expression in the hybrids, and reduced or absent expression in the parental (HONE1) cells and segregants. Through correlative analysis, we identified a small number of genes fitting this criteria, one of which mapped to the predefined critical region at 11q22.3-q23. The mRNA and protein expression patterns of this gene, called *THY1*, were then validated by RT-PCR and western blot analysis.

Examination of *THY1* protein levels in a tissue panel comprised of 9 normal nasopharyngeal mucosa specimens and 70 NPC samples revealed a substantially higher level of expression in the normal tissues, with the majority of NPC specimens showing reduced or absent *THY1* expression. By microscopic analysis, *THY1* protein was observed primarily in the cytoplasm and plasma membrane of the normal nasopharyngeal mucosa. The frequency of *THY1* reduction or absence in lymph node-metastatic NPC was 74% (20/27), which is significantly higher than that of primary NPC (40%), suggesting that the inactivation of *THY1* might also be associated with metastasis of NPC. Recently, *THY1* has been reported to play a role in the cell adhesion properties of T-cells adhering to the bone marrow stroma[69], and surface expression of *THY1* was shown to enhance focal adhesions in fibroblasts[70]. That an increasing body of evidence

now causally implicates loss of cell adhesion in tumor progression and metastasis[71] perhaps reflects an inroad for *THY1* into NPC pathogenesis.

THY1 is a cell surface glycoprotein of 25-28 kDa[72] expressed predominantly on the cytoplasmic membrane of T-cells, but has also been observed in the kidney, mammary gland, small intestine, skin and nervous system[73]. Functionally, *THY1* has been implicated as a trigger for several cellular processes including lymphokine release, proliferation, differentiation, and apoptosis[74-78]. However, the exact physiologic role of *THY1* remains unknown[79].

None of the tumorigenic NPC cell lines studies including HONE1, HK1, CNE1, HNE1 and the tumor segregants, were found to express *THY1* transcript or protein. Our data suggests that loss of *THY1* expression in NPC cell lines may be attributed to promoter hypermethylation, as at least one of the alleles was methylated in all four NPC cell lines, and demethylation by 5-aza-cytidine restored *THY1* expression in HONE1 cells.

Subsequent *in vitro* colony formation assays provided functional evidence that activation of wild type *THY1* is sufficient to inhibit the colony forming ability of *THY1*-null HONE1 cells using both the constitutive pCR3.1 and inducible pETE-BSD vector systems. The specificity of the growth suppressive effect of *THY1* was supported by the absence of suppression in the presence of dox, when the transgene was not expressed. These data thus functionally implicate *THY1* in NPC tumor suppression.

One discrepant observation in our study was the apparent lack of differential expression of *THY1* (by microarray) observed between HONE1 and the HK11.13 hybrid, and this hybrid and its tumor segregant (HK11.13-TS). By RT-PCR, we observed that *THY1* was undetectable in HONE1, HK11.13, and HK11.13-1TS after 45 PCR cycles. Such absence of expression could explain the lack of differential expression observed by microarray analysis; subsequent analysis confirmed that the microarray detection levels were substantially lower in these samples (data not shown). Genotyping results for this particular hybrid may provide a further explanation. The donor chromosome loci represented by microsatellite markers D11S908 and D11S2077, which map close to the *THY1* physical location, were observed to be absent in this particular hybrid[61], suggesting that the wild type *THY1* gene was either not transferred to this particular hybrid or was lost during or shortly after completion of the MMCT procedure. Nevertheless, while this explains the discrepant microarray and PCR observations, it fails to explain why *THY1* protein was presumably detected in this hybrid (but not in the parental or segregant cells) by western blot. Further genetic and biochemical analysis will be required to resolve this question.

Coincidentally, when searching the literature for clues to *THY1* function, we learned of a study whereby MMCT was used to introduce chromosome 11 into and ovarian cancer cell line, SKOV-3, and hybrids showing suppressed tumorigenicity were isolated from SCID mice. Using a subtractive hybridization technique, where cDNA populations from the tumorigenic parental line were subtracted from the nontumorigenic hybrids, the authors identified *THY1* as their top tumor suppressor candidate gene [85]. In a subsequent study, the authors showed that abrogation of *THY1* expression in non-tumorigenic cells could restore tumorigenesis; however, both *in vitro* and *in vivo* growth and tumorigenicity assays showed that *THY1* activation alone was not

sufficient to suppress ovarian cancer tumorigenicity[80], implying that other oncogenic mechanisms may interact with or modulate *THY1* tumor suppressor function.

In summary, our findings suggest that *THY1* is a novel tumor suppressor gene whose loss of expression by deletion or promoter methylation is a common event contributing to nasopharyngeal carcinogenesis. That *THY1* deficiency was significantly associated metastatic NPC suggests a possible role in NPC progression, and a clinical utility for *THY1* in patient prognosis. While further work is needed to characterize the precise molecular functions of THY in carcinogenesis, it should not be overlooked that a number of other genes mapping to various alternative chromosomal regions also matched the expression criteria of the theoretical tumor suppressor gene signature. How these genes may interact with *THY1* function, or contribute to tumor suppression independent of *THY1* function is an important question that we address in the next chapter of this thesis.

Chapter 3: Holistic Effects of MMCT: A Combinatorial Analysis of the NPC Altered Cancer Genome

3.0 Reconstructing the MMCT hypothesis

Mapping of tumor suppressor genes by MMCT-driven functional complementation involves the introduction of chromosomal material into a tumorigenic cell line, selection of nontumorigenic hybrids, and subsequent detailed chromosomal analysis of hybrid-derived tumor segregants. In this strategy, the tumor suppressor hypothesis is predicated on the assumptions: 1) that the introduced chromosome harbors one or more TSGs that act directly to suppress tumorigenicity in the hybrid, and 2) that the subsequent reversion to the tumorigenic phenotype (i.e., the segregant) results from loss of one or more TSGs on the introduced chromosome.

In this work, we investigate the ***alternative hypothesis***: that genes functionally contributing to the phenotypes of tumor suppression and reversion may stem from genomic alterations that occur ***outside*** of the introduced chromosome. Below, several plausible models aligned with this hypothesis are enumerated.

1. A gene present on the introduced chromosome could activate a TSG located on an endogenous chromosome (resulting in a nontumorigenic hybrid), and this endogenous chromosomal region could be subsequently *reduced in copy number* (resulting in a tumorigenic segregant).

2. A gene present on the introduced chromosome could inactivate an oncogene located on an endogenous chromosome (resulting in a nontumorigenic hybrid), and this endogenous

chromosomal region could be subsequently *increased in copy number* augmenting the activity of the oncogene (resulting in a tumorigenic segregant).

3. A TSG present on an endogenous chromosome could become activated by amplification during the MMCT procedure (resulting in a nontumorigenic hybrid), and this endogenous chromosomal region could be subsequently *reduced in copy number* (resulting in a tumorigenic segregant) with or without involvement of the introduced chromosome.

4. An oncogene present on an endogenous chromosome could become deactivated by copy number loss during the MMCT procedure (resulting in a nontumorigenic hybrid), and the same or different oncogene-harboring chromosomal region could be subsequently *amplified* (resulting in a tumorigenic segregant) with or without involvement of the introduced chromosome.

Several previous experimental observations also lend support to the possibility of the alternative hypothesis. First, while we identified *THY1* at 11q22.3 as a candidate tumor suppressor based on the original premise that an 11q critical region should harbor a TSG, we observed that one hybrid line (HK11.13) failed to express *THY1*. Microsatellite analysis suggested a genetic deletion of the gene prior to, or during, the MMCT procedure. This indicates that 1) unexpected chromosomal alterations can occur in the MMCT process, and 2) alternate tumor suppression mechanism may explain the "loss of tumorigenicity" phenotype. Furthermore, our whole-genome microarray expression analysis identified a number of genes residing on different chromosomes that also displayed the theoretical tumor suppressor signature of reduced expression in the hybrids, and increased expression in the parental and segregant lines (Chapter 2, Table T2.1).

3.1 Mining expression data for structural alterations

While by conventional MMCT it would not be possible to assess the comprehensive milieu of structural alterations in the parental, hybrid and segregant genomes, we postulated that a method of genome-wide locus-specific expression analysis may provide evidence for the existence of other recurrent chromosomal alterations that might explain the observed phenotypes. **This idea is based on the concept that genes within a given *amplified* chromosomal region would be significantly enriched for relative overexpression, while genes located within a *deleted* chromosomal region would be significantly enriched for relative underexpression.** To investigate this possibility, we developed a novel computational strategy involving the mapping of gene expression data to known chromosomal positions followed by density plot analysis based on the direction of gene expression (up or down) in parental-hybrid and hybrid-segregant comparisons. These plots were then used to visually and statistically identify candidate recurrent copy number alterations (CNAs) within the genome by determination of the most prominent density "peaks" and "troughs". Candidate CNAs resulting from this analysis were then validated using standard cytogenetic methods.

3.2 RESULTS

3.2.1 Data acquisition and processing

As shown in Chapter 2, Figure 2, we used an oligonucleotide microarray representing 19,000 genes to compare gene expression levels between: 1) HONE1 parental cells and each of four non-tumorigenic hybrids (11.8, 11.12, 11.13 and 11.19), and 2) each hybrid and its corresponding tumor segregant (11.18-3TS, 11.12-2TS, 11.13-1TS and 11.19-4TS). Dye-swap experiments (whereby two RNA samples are competitively hybridized to duplicate arrays, with the Cy3 and Cy5 sample labeling scheme being reversed between the two arrays) were carried out to control for dye labeling biases according to conventional practice (Appendix I 1.1). Normalized gene expression ratios were then log transformed, and the reciprocal ratios were taken from one array in the dye-swap pair and used to compute the average ratios for the two replicate experiments (Chapter 2, Table T2.2). Thus in practice, for each comparison, average expression ratios >1.0 were considered "overexpressed" in the hybrids, while ratios <1.0 were considered overexpressed in the HONE1 cells or segregants (depending on the comparison). Finally, as our aim was to uncover endogenous CNAs that might better explain the phenotypic observations than the chromosome-11-centric theory, we averaged the gene expression ratios across biological replicates (i.e., 4 replicates of parental (tumorigenic) versus hybrid (non-tumorigenic), and 4 replicates of hybrid versus tumorigenic segregant). This final averaging step allowed the gene expression data to be distilled down to a single quantitative measurement: >0 indicating increased expression in the hybrids, <0 indicating reduced expression in the hybrids (relative to HONE1 and segregant lines).

3.2.2 Characterizing genomic-expression density (GED) plots

To map the microarray expression data back to the genome, we traced the oligonucleotide probe sequences to GenBank accession numbers, and then mapped the GenBank sequences to the curated genome using the UCSC genome browser (http://genome.ucsc.edu/). Each gene was assigned two values: 1) the average expression value of the *parental versus hybrids* (PvH) comparison, and 2) the average expression value of the *hybrids versus segregants* (HvS) comparison. A sliding window of 5 MB size was initially selected for density plotting, as a window of this size was estimated to be sufficient for detecting known regions of deletion on chromosome 11 in the segregants. Genes within the window at a particular slide position were analyzed for the enrichment of up-regulated or down-regulated gene expression. By convention, in our *genomic-expression density* (GED) plots the enrichment of down-regulated genes (in the parental or segregant lines) corresponded to a high score (near to 1.0); equal distribution of up- and down-regulated genes corresponded to a "baseline" value of ~0.5, and enrichment of up-regulated genes corresponded to a score near to 0. All chromosomes were scanned using the 5 MB window sliding at one-gene intervals, and peaks that achieved values above 0.7, or troughs that reached values below 0.3, were chosen for statistical analysis by Fisher's exact test to determine the significance of the distribution of up- or down-regulated genes in that region.

Figure F3.1A shows a representative genomic-expression density (GED) plot for chromosome 15, which lacks evidence of copy number changes. Neither the PvH (in orange) nor the HvS (in green) GED plots showed enrichment fluctuations above 0.7 or below 0.3. For chromosome 11, however, several interesting regions emerged. Two distinct regions at 11q13.4-14.1 and 11q21-23.3 were found to be enriched for up-regulated gene expression in the hybrids

(PvH plot) and down-regulated gene expression in the segregants (HvS plot) consistent with a "phenotypically-balanced" event (i.e., a copy number gain in the hybrid, with subsequent copy number loss in the segregant) (Figure F3.1B). Statistical analysis suggested that these observations would, individually, be unlikely to occur by chance (11q13.4-14.1, P=0.0008 [PvH], P=0.006 [HvS]; 11q21-23.3, P=0.000009 [PvH], P=0.0002 [HvS]). Notably, the previously described *THY1* cell surface antigen gene we identified as a novel tumor suppressor in NPC is located in the 11q22.3 region (Chapter 2). Both microsatellite and FISH analysis of 11q13.4-14.1 and 11q21-23.3 in the segregant cell lines previously established a consistent loss of heterozygosity in both regions [67]. Thus, the GED plot analysis correctly identified two chromosomal regions known to display copy number gain in the hybrids and subsequent copy number loss in the segregants.

The 11p15.1-14.1 locus was also identified here as a candidate region of copy number loss in the segregants (P=0.002; Figure F3.1B). Interestingly, while this region has not been previously analyzed by microsatellite or FISH analysis in our NPC segregants, it has been implicated as a tumor suppressor locus elsewhere, showing loss of heterozygosity (LOH) in several adult and childhood cancers [81].

A detailed GED plot analysis of the whole genome revealed evidence for several other phenotypically-balanced recurrent alterations. In chromosome 1, two adjacent loci located at 1q23.3-23.5 and 1q32.1-31.2 showed evidence of copy number gain in the hybrids (P=0.002, P=0.002, respectively) and loss in the segregants (P=5.5x10^{-10}, P=1.8x10^{-8}, respectively) (Figure F3.2A). Interestingly, loss of chromosome 1q23 was recently reported to play a role in the

initiation and progression of insulinomas and esophageal squamous cell carcinomas [82]. In chromosome 16, two interesting loci located at 16q12.2 and 16p13.3-13.2 were identified as candidate recurrent CNAs (Figure F3.2B). The 16q12.2 region appeared to only be amplified in the segregants (P=0.0006), while 16p13.3-13.2 appeared more "phenotypically balanced" with putative copy number loss in the hybrids and concomitant amplification in the segregants (P=1.3x10^{-5} [PvH], P=8.7x10^{-6} [HvS]). Notably, in a recent study of breast cancer copy number alterations, 16p13 was observed to be amplified in over half of the tumors studied [83]. The phenotypically-balanced candidate CNAs were analyzed further using molecular and cytogenetic approaches.

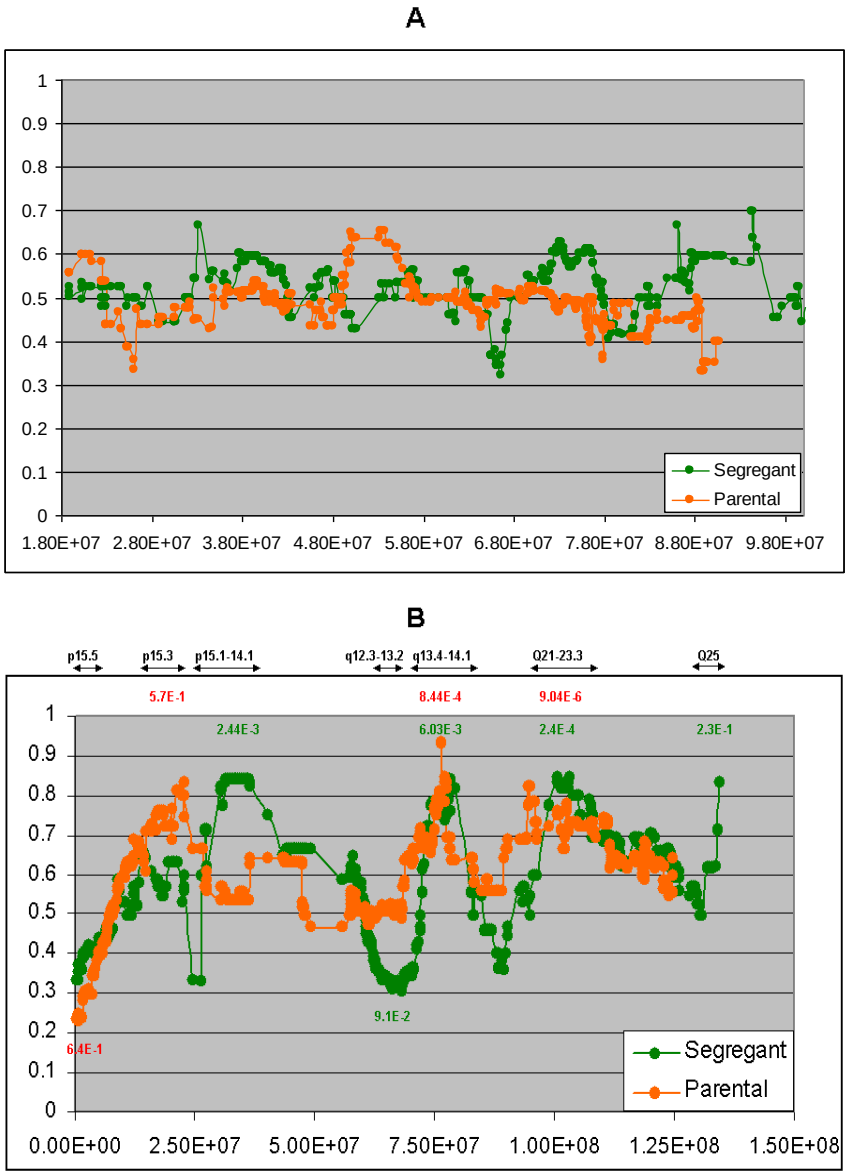


Figure F3.1: Genomic-expression density plots.

A) Representative GED plot of the averaged expression data for chromosome 15.
B) GED plot of Chromosome 11. Peaks (>0.7) and troughs (<0.3) represent loci where genes are disproportionately down- or up-regulated. Green peaks reflect local down-regulated expression in the segregants (relative to hybrids), and orange peaks represent local down-regulated expression in the parental cells (relative to hybrids).) The statistical significance of peaks and troughs was calculated using Fisher's exact test; p-values are shown in corresponding color above peaks and below troughs. Chromosomal bands of interest are shown at the top.

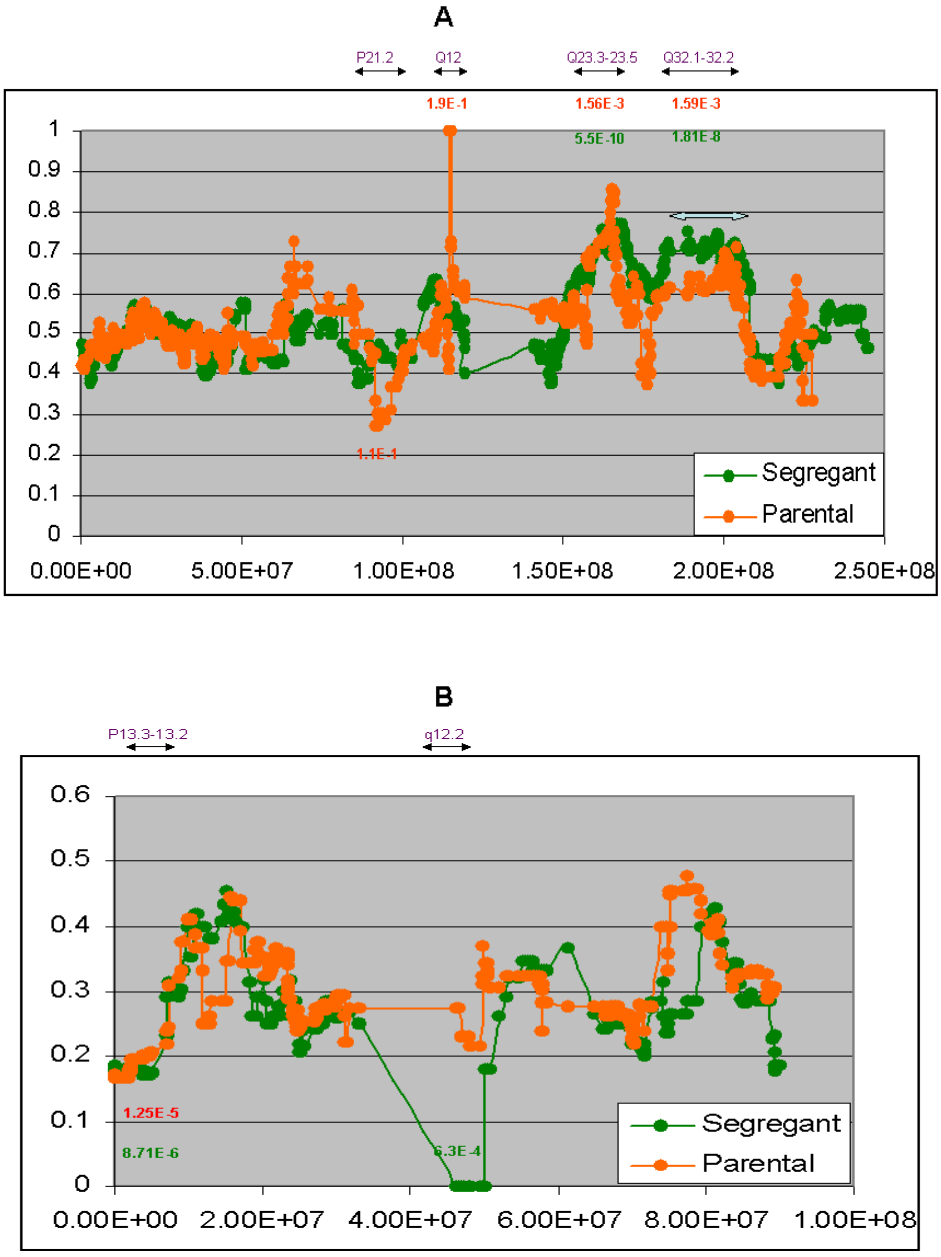


Figure F3.2: GED plots of (A) Chromosome 1 and (B) Chromosome 16.

Peaks (>0.7) and troughs (<0.3) represent loci where genes are disproportionately down- or up-regulated. Green peaks reflect local down-regulated expression in the segregants (relative to hybrids), and orange peaks represent local down-regulated expression in the parental cells (relative to hybrids).) The statistical significance of peaks and troughs was calculated using Fisher's exact test; p-values are shown in corresponding color above peaks and below troughs. Chromosomal bands of interest are shown at the top.

3.2.3 Confirmation by Quantitative real-time PCR

Quantitative real-time (RT) PCR was used to confirm the microarray/GED plot findings of locus-specific differences in gene expression between hybrids and their respective segregants (Appendix I 1.12). At each locus, a panel of genes was analyzed for expression differences. Analysis of genes located in both the 1q23.3-23.5 region and the 1q32.1-31.2 locus showed a considerable and consistent decrease in transcript levels in the segregants (Figure F3.3A & 3.3B), consistent with the GED plot data. Conversely, RT PCR analysis of genes at chromosome 16p13.3-13.2 showed a substantial and consistent increase in transcript levels in the segregants (Figure F3.4). Thus, quantitative PCR analysis of gene transcript levels corroborates the microarray expression data and the GED plot findings that suggest locus-specific copy number alterations.

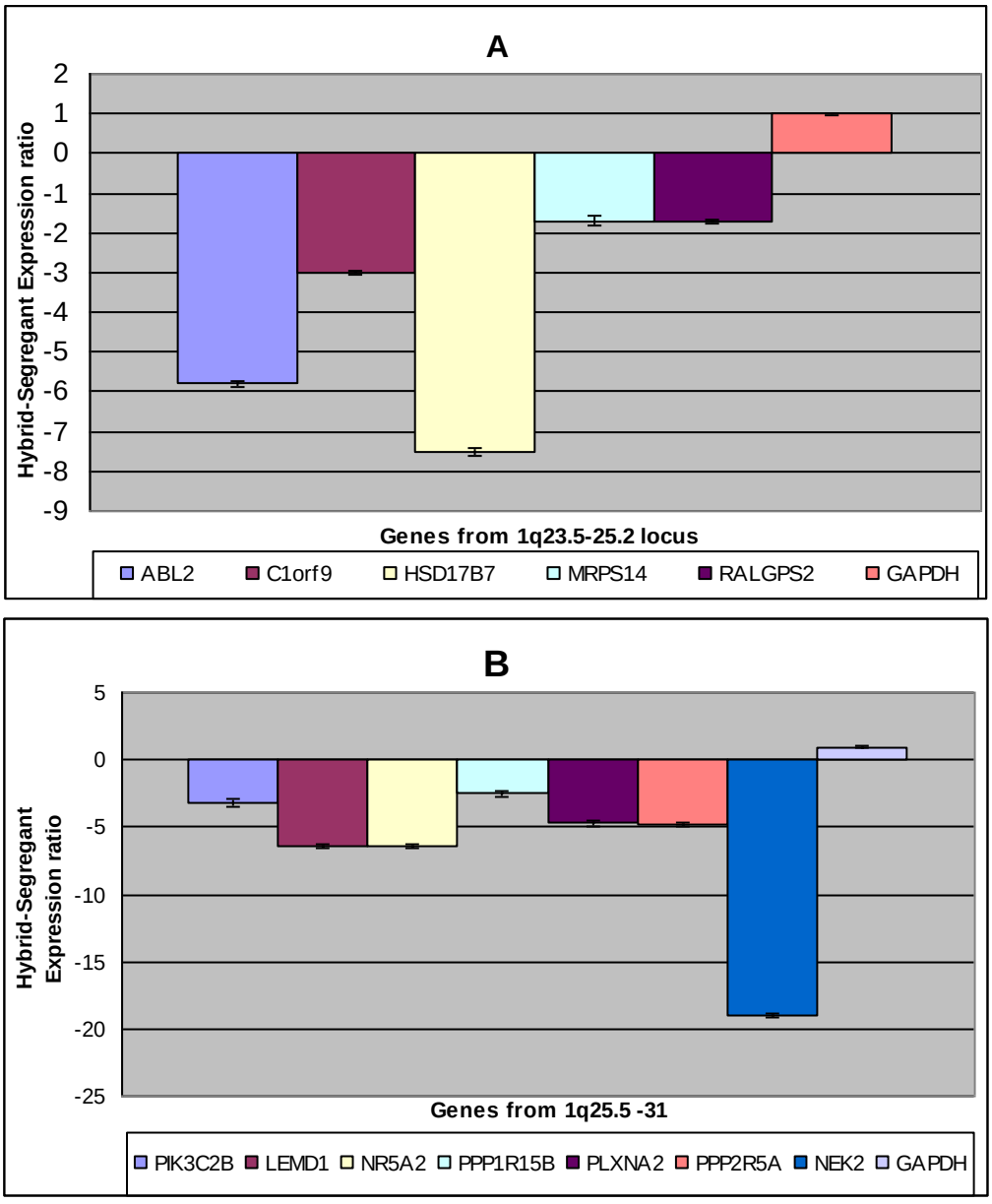


Figure F3.3: Quantitative PCR of the genes in Chromosome 1 (11.8 hybrid).

A) Genes from the 1q23.3-23.5 locus and
B) genes from the 1q32.1-31.2 locus are shown. The histogram shows the average fold difference between the hybrids and the segregants. GAPDH was used as internal control.

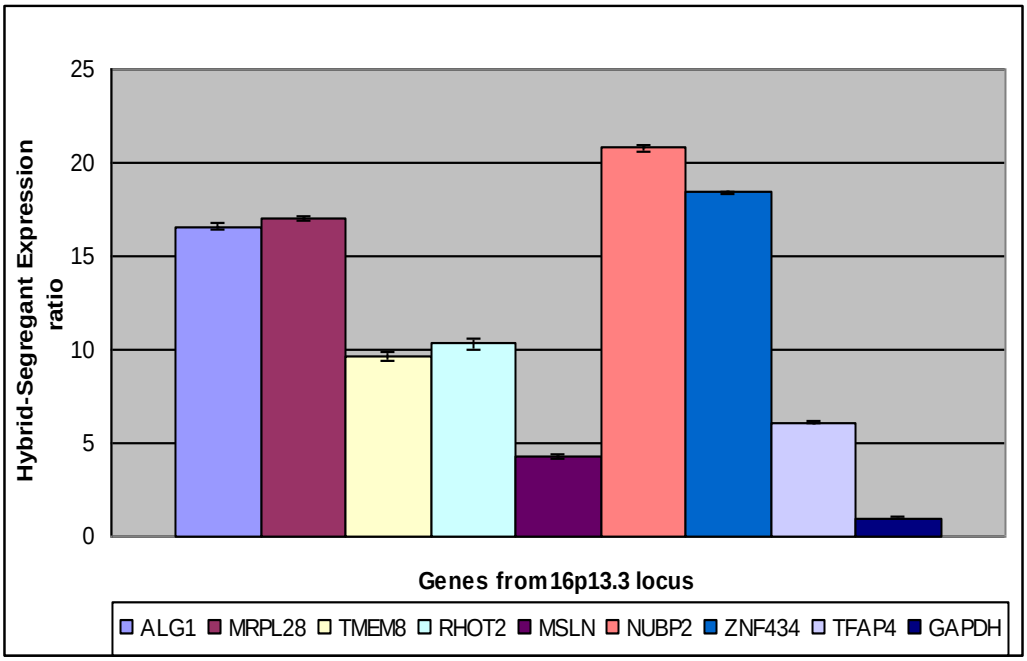


Figure F3.4: Quantitative PCR of the genes of Chromosome 16p13.3-13.2 in 11.8.

The Histogram shows the average fold difference between hybrids and segregants.

3.2.4 Cytogenetic validation of the GED plot analysis

Two of the four hybrid-segregant pairs of cell lines were available to investigate the accuracy of the whole-genome GED plot analysis in uncovering novel recurrent CNAs. Using FISH and SKY methodologies, several candidate regions identified by the GED plot were investigated. FISH analysis was performed using a panel of probes mapping to the regions of interest (Table T3.1) (Appendix I 1.13 & 1.14). As described below, for both of the hybrid-segregant pairs, the cytogenetic results were highly consistent with the GED plot observations. At chromosome 1q23.3-23.5 and 1q32.1-31.2, 5/7 probes and 4/6 probes, respectively, showed a clear and consistent loss of either one or two copies in the segregants (relative to the hybrid) (Table T3.1 and Figure F3.5). Those two probes from each region did not show a copy number change indicates that the loss of heterozygosity may not be contiguous (Table T3.1).

Similarly, chromosome position 16p13.3-13.2 showed an increase in copy number in the segregant (relative to the hybrid) (Figure F3.5). Along the chromosome at 1q23.3-23.5 most of the probes showed consistent loss of either one or two copies and intermittently some of the probes didn’t gain or lose any copy number indicating again that the loss is not completely contiguous. The chromosome 16p13.3-13.2 region is relatively small, and all the probes that were tested showed the same increase in copy number (Table T3.1). SKY analysis of the hybrids consistently revealed the presence of nearly 7 copies of chromosome 1, out of which only three were found to be intact, while the rest were truncated and often fused with fragments from other chromosomes (Figure F3.5, lower panels). In the segregant lines, SKY analysis revealed a consistent loss of one or two copies of chromosome 1. Chromosome 16, on the other hand, showed four copy numbers in both hybrids and segregants, with the exception of a fragment of

chromosome 16 found translocated to chromosome 8 only in the segregants. This translocated piece could account for the consistent single copy gain in the segregants. Taken together, these observations suggest that the GED plot analysis, which uses only transcriptional information to measure genome-localized expression fluctuations (i.e., consistent with copy number gains and losses), can in fact correctly infer genomic copy number changes of subtle magnitude.

A

Probes(Chr1q23.3)	Hybrid	Segregant	Loss	Start	Stop
RP11-79M15	6	6	0	157484198	157580781
RP11-154F13	6	4	2	165964956	166136565
CTD-2171O24	6	4	2	166048311	166122915
CTD-2166B20	6	4	2	167346697	167442066
CTD-3226D17	6	6	0	159490628	159517234
RP11-18E13	6	5	1	175377432	175553514
RP11-152A16	6	4	2	175804858	175986939

B

Probes(Chr1q32.1)	Hybrid	Segregant	Loss	Start	Stop
CTD-2501I18	6	4	2	9057132	9208284
CTD-2332F8	6	6	0	1.97E+08	1.97E+08
RP11-284G5	6	4	2	2.01E+08	2.01E+08
CTD-2288C12	6	6	0	2.02E+08	2.02E+08
RP11-243A2	6	5	1	2.05E+08	2.05E+08
RP11-22D13	6	5	1	2.09E+08	2.09E+08

C

Probes(Chr16p13.3)	Hybrid	Segregant	Gain	Start	Stop
CTD-2252G4	4	5	1	11227	131219
RP11-243K18	4	5	1	266994	431803
RP11-64L12	4	5	1	535529	715973
RP11-31I10	4	5	1	1752997	1918786
RP11-626F12	4	5	1	2825861	3016463
RP11-95P2	4	5	1	4114617	4289638
CTD-3103H12	4	5	1	4488744	4545592
RP11-113H7	4	5	1	5041344	5214907

Table T3.1: List of FISH probes, their copy number status in hybrids and segregants, and their position in the genome are provided.

A) Probes from chromosome 1q23.3-23.5 locus.
B) Probes from chromosome 1q32.1-31.2. **C)** Probes from chromosome 16p13.3-13.2 locus.

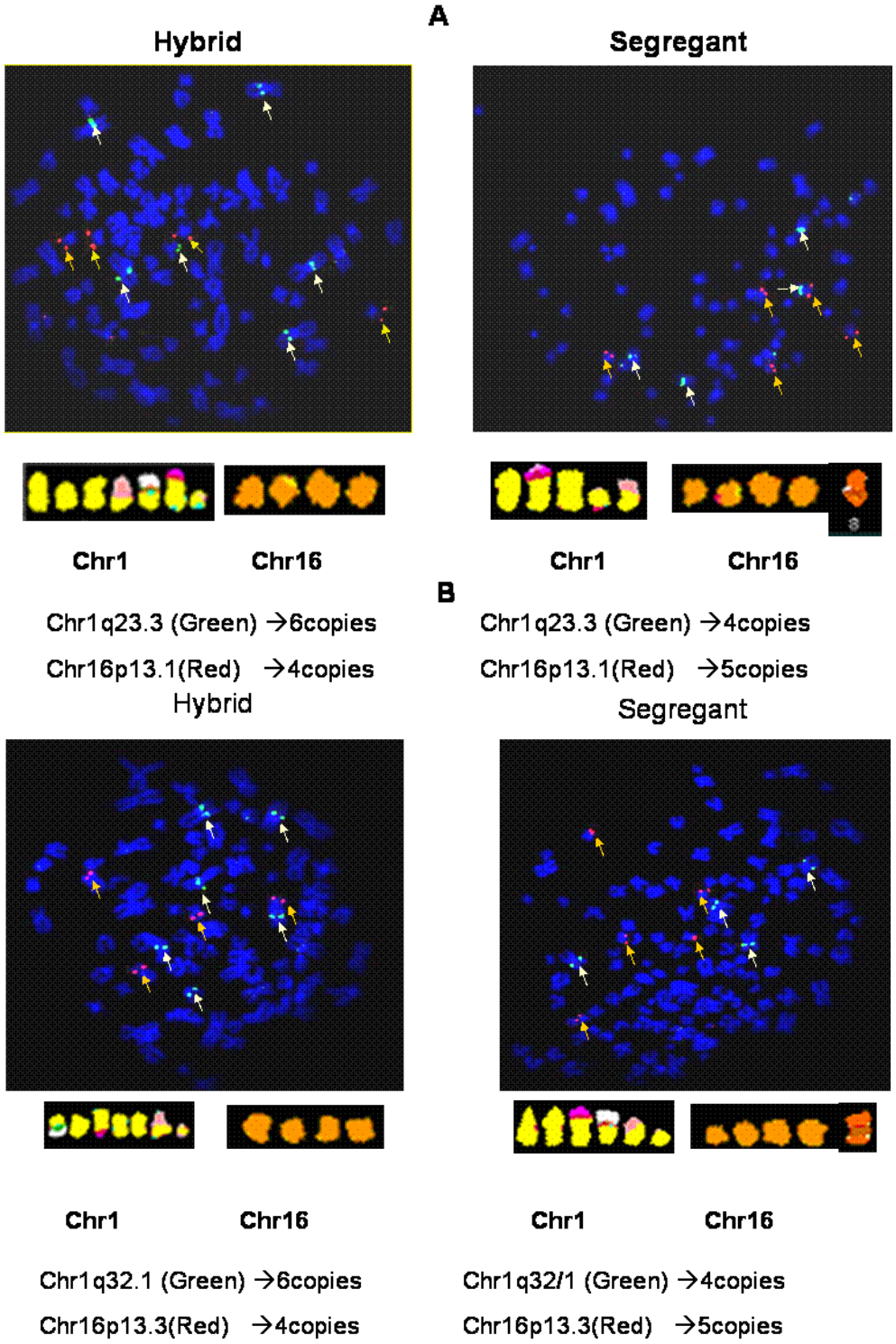


Figure F3.5: FISH and SKY results of NPC hybrid and segregant lines.

White arrows indicate the signals in chromosome 1; yellow arrows mark the signals in chromosome 16. SKY images are shown below the FISH results.
A) hybrid (11.8) and segregant names (11.8_ts) and **B)** hybrid (11.19) and segregant names (11.19_ts).

3.3 Discussion

In this work, we have challenged the age-old assumption that TSG discovery by MMCT-based functional complementation should focus only on genes located within the exogenously introduced chromosome. By conducting a systematic genomic analysis of locus-specific differential gene expression between HONE1, hybrid, and segregant cells (i.e., by GED plot analysis), we have elucidated a number of candidate recurrent copy number alterations that we suspect harbor one or more genes contributing to the observed tumor phenotypes. Known alterations on chromosome 11 served as a positive control in our model, demonstrating that genomic-expression density plot analysis can accurately discover underlying chromosomal imbalances distinguished by locus-specific imbalances in gene expression.

Several of the phenotypically-balanced candidate loci were examined in detail by quantitative real time PCR as well as the cytogenetic methods, FISH and SKY. Two adjacent regions on chromosome 1, 1q23.3-23.5 and 1q32.1-31.2, showed evidence of copy number gain in two hybrids and subsequent copy number loss in the two corresponding segregant lines, while the p13.3-13.2 locus on chromosome 16 showed evidence of copy number loss in the hybrids followed by copy number gain in the corresponding segregants. Due to political constraints encountered by our collaborators, we have thus far only been able to validate several of our candidate CNAs, in only two of the four hybrid-segregant pairs (and not including the original HONE1 parental line). However, for future publication of this work, this problem will be rectified; we are already preparing to extend the cytogenetic analysis of these loci to the remaining hybrid-segregant pairs and the original HONE1 parental cells. Furthermore, we also

will examine cytogenetically the 11p15 region suspected to be deleted in the segregants, as well as several other candidate regions of phenotypically-*unbalanced* copy number gain.

HONE1 nasopharyngeal carcinoma cells are hypo-tetraploid with a large number of both segmental and whole-chromosome imbalances. As shown in Fig. 5, the copy number losses at 1q and copy number gain at 16p (observed in the segregants) are more representative of subtle dosage effects, rather than homozygous allelic deletion or high-level amplification. Nevertheless, their detection by *averaged* expression levels across four biological replicates (parental to hybrid, and hybrid to segregant) and their cytogenetic validation in two of the replicates, indicate that these alterations, though subtle, are recurrent in our experimental model and thus consistent with a role in mediating the observed tumorigenic phenotypes.

By RT-PCR analysis, we examined the impact of the CNAs on local gene expression. We found that the average gain in gene expression at 16p13.3-13.2 was approximately 13-fold higher in the segregants relative to the hybrids. Conversely, at 1q23.3-23.5, the average reduction of gene expression in the segregants was roughly 4-fold lower relative to the hybrids. Thus, while the copy number differences appeared subtle, the impact on gene expression was more substantial, with fold differences that we postulate could manifest in the observed phenotypic differences.

The alterations discovered at 1q best fit model #3 of the alternative hypothesis described in the first section of this chapter: that a TSG present on an endogenous chromosome could become activated by amplification during the MMCT procedure (resulting in a nontumorigenic

hybrid), and this endogenous chromosomal region could be subsequently *reduced in copy number* (resulting in a tumorigenic segregant) with or without involvement of the introduced chromosome. In this case, our data suggest that endogenous TSG(s) reside on the identified 1q regions, and that the hybrid cells gain copies of these regions during chromosomal transfer/hybrid generation by either amplification or clonal selection. The clonal selection theory can be explained by a subpopulation of HONE1 cells characterized by increased copy number at the 1q regions and thus, presumably associated with enhanced potential for TSG activation. In the presence of exogenous chromosome 11, this subpopulation may have a selective advantage for giving rise to a nontumorigenic phenotype. Reciprocally, the TSG activity would become suppressed as the hybrids lose copies of these 1q regions during progression towards the segregant phenotype. In the case of 16p, a reduction in copy number was associated with hybrid formation, while a gain in copy number was associated with progression to a segregant phenotype. This phenomenon is most related to model #4, whereby an oncogene present on an endogenous chromosome could become deactivated by copy number loss during the MMCT procedure (resulting in a nontumorigenic hybrid), and the same oncogene-harboring chromosomal region could be subsequently *amplified,* augmenting the activity of the oncogene, (resulting in a tumorigenic segregant). Again, the clonal selection theory may be invoked. In this case, a HONE1 subpopulation with reduced copy number of a 16p13.3-13.2-localized oncogene (and thus suppressed oncogenic activity) could selectively give rise to a nontumorigenic hybrid. Hybrid clones could then progress towards the segregant phenotype by reactivation of the oncogene through selective amplification.

An interesting question is whether one chromosomal locus is alone sufficient to drive the tumor suppression (hybrid) and reversion (segregant) phenotypes. Previously, we identified THY1 at 11q22.3 as a candidate tumor suppressor gene in our model of NPC tumorigenicity. We observed that THY1 expression could modulate the growth properties of NPC cells in a manner consistent with a role in tumor suppression. However, that multiple other chromosomal changes were repeatedly selected for in our "parental-hybrid-segregant" replicates, suggests that each recurrent locus, together with THY1, plays a necessary role in a larger and more complex tumorigenic program than can be explained by the activity of a single gene.

In conclusion, our results suggest that GED plot analysis can shed light on previously unanticipated recurrent copy number changes in endogenous chromosomes that may contribute to the tumor suppression/reversion phenotypes. While our data implicate a small number of specific genomic modifications, they neither tell us which specific genes or pathways are being selected for, nor how they interact with elements of chromosome 11. In the following chapter, we seek to better understand the biology behind our observations by building on the copy number-gene expression intersect with integration of a third tier of information: pathway enrichment.

Chapter 4: Oncogenomics and Pathway Discovery in NPC Progression

4.0 From genomic alterations to signaling pathways

In the previous chapters, we have used an MMCT-based functional complementation approach to model how tumorigenic properties of NPC cells are lost and gained from a genomic perspective. While the introduction of chromosome 11 conferred a nontumorigenic phenotype to hybrid cells, and specific losses of chromosome 11 were associated with reversion to the tumorigenic phenotype, other phenotype-correlated recurrent CNAs were also observed, implying the existence of a complex interplay between multiple genes and tumorigenicity. In this work, we sought to gain insight into NPC tumorigenicity from a pathway perspective, rather than a purely genomic one. We revisited the question of a "tumor suppressor signature" presented in Chapter 2, where we identified genes whose expression patterns varied according to phenotypic state. Genes up-regulated in the hybrids (relative to parental) and down-regulated in the segregants (relative to hybrids) were of primary interest. Here, we also considered genes with the reciprocal pattern (up in parental cells relative to hybrids, and down in hybrids relative to segregants) for a comprehensive assessment of all differentially expressed genes in our model of tumorigenicity.

In the following chapter, we describe in detail how the application of a statistical gene ontology analysis to sets of differentially expressed genes revealed the involvement of multiple signaling pathways, and further explore by clustering, promoter mining, and biochemical validation the specific involvement of sterol biosynthesis in the progression of NPC.

4.1 RESULTS

4.1.1 Ontology and pathway analyses

From the microarray experiments, differentially expressed genes were selected for each of the two phenotypic comparisons: Parental versus Hybrid (***PvH***; 827 genes) and Hybrid versus Segregant (***HvS***; 714 genes). Genes that showed at least a 1.5 fold change consistently, across all replicate cell-line comparisons and in the same direction, were considered to be differentially expressed. The differentially expressed genes were then categorized according to their direction of expression (i.e., up- or down-regulated) for gene ontology analysis (Table T4.1).

Comparison	**Total # diff. exp.**	**Up-regulated in parental/segregant**	**Down-regulated in parental/segregant**
Parental vs. Hybrid	827	442	385
Hybrid vs. Segregant	714	387	327

Table T4.1: Differentially expressed genes in PvH and HvS comparisons.

Pathway and gene ontology analyses were carried out separately on the gene expression categories to identify significantly enriched functional classes of genes. We used the **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (DAVID) tool from the National Institute of Allergy and Infectious Diseases (NIAID)[84], NIH for the functional annotation analysis. DAVID contains a knowledgebase that integrates and conglomerates the major and well-known public bioinformatics resources such as NCBI, PIR, SWISS-PROT, GO, OMIM, PubMed, KEGG, BIOCARTA, Affymetrix, TIGR, Pfam, BIND, MINT and DIP using a graph theory evidence-based "DAVID Gene Concept". The knowledge base is the central engine of this tool on which several analytical domains are based. The "gene enrichment and functional annotation

analysis" domain of DAVID was used to analyze the gene lists for statistical enrichments of GO terms and KEGG and BIOCARTA pathways. A modified Fisher Exact score (more stringent than the classical Fisher exact score) termed "EASE score" was used to statistically identify the significant subsets of genes associated with a particular functional term or pathway (See Appendix I 1.15)[85]. This analysis revealed a number of statistically significant functional gene enrichments within each of the gene lists examined. The results are shown in Tables T4.2-T4.5.

First, we observed enrichment of a number of known cancer-related processes that corresponded well with the tumorigenic phenotype [86]. For example, in the HvS comparison, *Cell Adhesion*, *Cell Motility*, *Angiogenesis*, and *Blood Vessel Development* were all significantly enriched in the genes up-regulated in the tumor segregants; while *Cell Adhesion*, *Cell Death*, and *Apoptosis* were all enriched in the genes down-regulated in the segregants. Second, we observed that some enriched terms were "phenotypically balanced", eg, down-regulated in parental (PvH) and also down-regulated in the segregants (HvS) – "balanced" with respect to reduced expression in the tumorigenic phenotypes (parental and segregants). For example, just as *Cell Adhesion*, *Cell Death*, and *Apoptosis* were found enriched in the list of genes down-regulated in the segregants (relative to hybrids), these same processes were also enriched for in the down-regulated genes of the parental cells (relative to hybrids).

One of the most interesting and unexpected phenotypically-balanced findings involved steroid metabolism/biosynthesis genes. GO analyses involving the HvS down-regulated genes (down in segregants) revealed statistically significant Ease scores of $P = 9.0E\text{-}13$ and $P = 3.67E\text{-}10$, respectively, for *Steroid Biosynthesis* and *Steroid Metabolism*; while Pathway analysis

resulted in an Ease score of *P = 7.62E-9* for *Biosynthesis of Steroids* (Table T4.2). Reciprocally, GO analyses of the PvH down-regulated genes (down in parental) showed statistically significant enrichment for *Steroid Biosynthesis* and *Steroid Metabolism* (*P = 6.0E-4* and *P = 1.4E-2*, respectively); as well as Pathway enrichment of *Biosynthesis of Steroids* (*P = 4.66E-5*) (Table T4.3). Consistent with these observations, the list of 24 top candidate (tumor suppressor) genes derived using the "Pearson correlation coefficient" score-based ranking consisted of seven (i.e., 29%) key steroid metabolism regulatory genes (Figure F4.1). Taken together, these observations demonstrate a significant correlation between reduced expression of steroid metabolism/synthesis genes and tumorigenic growth, implicating the biology of steroid production in the maintenance of a nontumorigenic phenotype.

One subgroup of steroids, the sterols, are amphipathic lipids synthesized from acetyl-coenzyme A – and the predominant sterol in vertebrate cells is cholesterol. Like *Steroid Biosynthesis*, *Lipid Biosynthesis* was also significantly enriched in the down-regulated genes of both the HvS and PvH comparisons (*P = 9.24E-8* and *P = 8.15E-3*, respectively); while *Cholesterol Biosynthesis*, specifically, had an Ease score of *P = 7.19E-7* (the 4th most significant score) in the down-regulated genes of the HvS comparison. Thus, one interpretation of these observations is that, more specific than general steroid production, the production of sterols, and perhaps cholesterol, in particular, may be involved in tumor suppression.

At the pathway level, *TNFR2 Signaling Pathway* was the second most significant pathway in the HvS down regulated gene list (*9.69E-3*) (Table T4.2 & T4.3). Interestingly, the

activation of this pathway is known to be associated with both tumor growth and tumor death, and as such may play a functional role in our model of NPC.

Down Regulated in Segregants

GENE ONTOLOGY

Term	Count	%	P Value
STEROID BIOSYNTHESIS	13	14%	9.0E-13
STEROID METABOLISM	14	15%	3.67E-10
LIPID BIOSYNTHESIS	14	15%	9.24E-8
CHOLESTEROL BIOSYNTHESIS	7	7%	7.19E-7
ALCOHOL METABOLISM	10	10%	2.96E-6
CHOLESTEROL METABOLISM	7	7%	1.80E-4
CELL DEATH	7	7%	3.41E-2
DEATH	7	7%	3.55E-2
CELL MIGRATION	3	3%	4.34E-2
ORGANOGENESIS	10	10%	4.38E-2
AMINE METABOLISM	5	5%	6.12E-2
REGULATION OF CELLULAR PHYSIOLOGICAL PROCESS	7	7%	7.50E-2
APOPTOSIS	6	6%	7.67E-2
CELL ADHESION	7	7%	8.11E-2
AROMATIC COMPOUND METABOLISM	3	3%	8.76E-2
POSITIVE REGULATION OF PHYSIOLOGICAL PROCESS	5	5%	9.10E-2
CELLULAR PHYSIOLOGICAL PROCESS	57	61%	9.88E-2

KEGG PATHWAY

BIOSYNTHESIS OF STEROIDS(Homo sapiens)	7	7%	7.62E-9
TERPENOID BIOSYNTHESIS(Homo sapiens)	2	2%	5.3E-2

BIOCARTA

TNFR2 Signaling Pathway(Homo sapiens)	3	3%	9.69E-3

Table T4.2: Gene ontology and pathway analysis of down regulated genes in the segregants showing the significantly enriched pathways and GO terms.

Up Regulated in Segregants

GENE ONTOLOGY

Term	Count	%	P Value
CELL ADHESION	9	10%	3.93E-3
DEVELOPMENT	16	18%	7.35E-3
CELL MOTILITY	5	5%	9.17E-3
ORGANOGENESIS	10	11%	1.73E-2
ANGIOGENESIS	3	3%	2.05E-2
BLOOD VESSEL DEVELOPMENT	3	3%	2.27E-2
CELLULAR PHYSIOLOGICAL PROCESS	51	60%	3.34E-2
TRNA PROCESSING	3	3%	4.43E-2
REGULATION OF BIOLOGICAL PROCESS	20	23%	4.46E-2
IMMUNE CELL MIGRATION	2	2%	5.19E-2
POSITIVE REGULATION OF PHYSIOLOGICAL PROCESS	5	5%	5.7E-2

KEGG PATHWAY

RIBOSOME(Homo sapiens)	13	2%	1.53E-5
TGF-BETA SIGNALING PATHWAY(Homo sapiens)	7	1%	6.18E-2
ARGININE AND PROLINE METABOLISM(Homo sapiens)	5	0%	8.79E-2
TRYPTOPHAN METABOLISM(Homo sapiens)	7	1%	9.45E-2

BIOCARTA

Melanocyte Development and Pigmentation Pathway(Homo sapiens)	3	0%	4.44E-2
Transcription Regulation by Methyltransferase of CARM1(Homo sapiens)	3	0%	4.44E-2

Table T4.3: Gene ontology and pathway analysis of up regulated genes in the segregants showing the significantly enriched pathways and GO terms.

Down Regulated in Parental

GENE ONTOLOGY

Term	Count	%	P Value
CELL-CELL SIGNALING	19	8%	5.82E-5
CELL DEATH	17	7%	1.83E-4
DEATH	17	7%	2.5E-4
MAINTENANCE OF PROTEIN LOCALIZATION	4	1%	3.65E-4
ALCOHOL METABOLISM	11	4%	5.27E-4
REGULATION OF CELLULAR PHYSIOLOGICAL PROCESS	18	7%	5.72E-4
STEROID BIOSYNTHESIS	8	3%	6.0E-4
MAINTENANCE OF LOCALIZATION	4	1%	6.98E-4
APOPTOSIS	15	6%	9.8E-4
CELL COMMUNICATION	64	27%	9.29E-4
CELL PROLIFERATION	26	11%	1.55E-3
CELL ADHESION	17	7%	1.93E-3
RESPONSE TO EXTERNAL STIMULUS	28	11%	3.36E-3
NEGATIVE REGULATION OF PHYSIOLOGICAL PROCESS	13	5%	5.40E-3
NEGATIVE REGULATION OF CELLULAR PHYSIOLOGICAL PROCESS	10	4%	5.89E-3
LIPID BIOSYNTHESIS	12	5%	8.15E-3
STEROID METABOLISM	9	3%	1.4E-2

KEGG PATHWAY

BIOSYNTHESIS OF STEROIDS(Homo sapiens)	6	2%	4.66E-5
TERPENOID BIOSYNTHESIS(Homo sapiens)	4	1%	1.79E-4
CALCIUM SIGNALING PATHWAY(Homo sapiens)	10	4%	1.67E-2
CIRCADIAN RHYTHM(Homo sapiens)	3	1%	3.96E-2

Table T4.4: Gene ontology and pathway analysis of down regulated genes in the parental cell lines showing the significantly enriched pathways and GO terms.

Up Regulated in Parental

GENE ONTOLOGY

Term	Count	%	P Value
CELLULAR PHYSIOLOGICAL PROCESS	345	58%	9.71E-13
NUCLEOBASE, NUCLEOSIDE, NUCLEOTIDE AND NUCLEIC ACID METABOLISM	136	23%	1.33E-7
CELLULAR METABOLISM	255	43%	1.24E-6
REGULATION OF BIOLOGICAL PROCESS	123	20%	6.19E-6
TRANSCRIPTION	87	14%	1.93E-4
CELLULAR BIOSYNTHESIS	50	8%	2.44E-4
REGULATION OF METABOLISM	88	14%	2.93E-4
PROTEIN METABOLISM	113	19%	3.13E-4
RIBOSOME BIOGENESIS	8	1%	3.47E-4
CELLULAR PROTEIN METABOLISM	112	19%	3.53E-4
DEVELOPMENT	74	12%	5.81E-4
RIBOSOME BIOGENESIS AND ASSEMBLY	8	1%	7.03E-4
MACROMOLECULE METABOLISM	121	20%	1.21E-3
NEGATIVE REGULATION OF METABOLISM	13	2%	2.22E-3
CYTOPLASM ORGANIZATION AND BIOGENESIS	8	1%	2.64E-3
NEGATIVE REGULATION OF PHYSIOLOGICAL PROCESS	24	4%	2.84E-3
ORGANOGENESIS	42	7%	2.93E-3
ORGANELLE ORGANIZATION AND BIOGENESIS	32	5%	3.94E-3
RNA PROCESSING	33	5%	4.88E-3

KEGG PATHWAY

Pyrimidine Metabolism	29	1.4%	1.0E-6
Cell cycle	12	0.6%	1.0E-4
RNA polymerase	34	1.7%	2.8E-4
Purine metabolism	15	0.7%	1.2E-3

Table T4.5: Gene ontology and pathway analysis of up regulated genes in the segregants showing the significantly enriched pathways and GO terms.

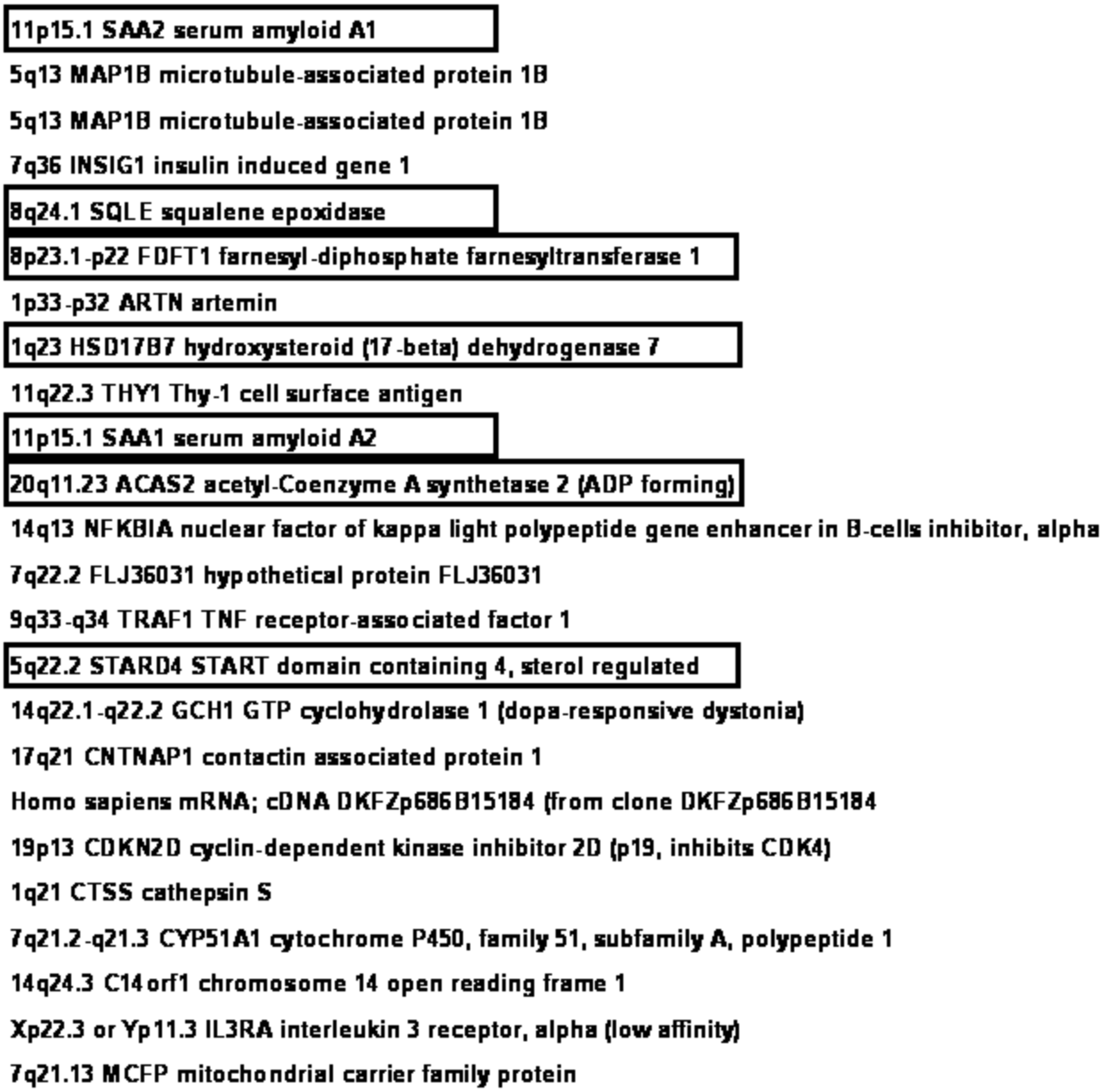


Figure F4.1: Top 24 candidate tumor suppressor genes.

Genes were ranked according to Pearson correlation coefficient scores (see Chapter 2, Table 1). The sterol regulatory genes are shown enclosed by boxes.

4.1.2 PCR verification of differentially expressed sterol and TNFR2 signaling pathway genes

Quantitative PCR analysis (Appendix I Section 1.12) was carried out to validate the differential expression of genes associated with sterol and TNFR2 signaling pathways. SAA1, SAA2, INSIG1 and INSIG2 function as key regulators of cholesterol transport and intracellular concentration [87, 88]. TRAF1, A20 (TNFAIP3) and IKB-α are essential downstream components of the TNFR2 signaling pathway. By microarray analysis, all seven genes were observed to be consistently down-regulated in the segregants in the HvS comparisons. As shown in Figures F4.2A and F4.2B, quantitative PCR confirmed the microarray results, showing an average of 13-fold repression (9-fold to 18-fold) in the sterol pathway genes, and an average repression of 16-fold (9-fold to 23-fold) in the TNFR2 pathway genes.

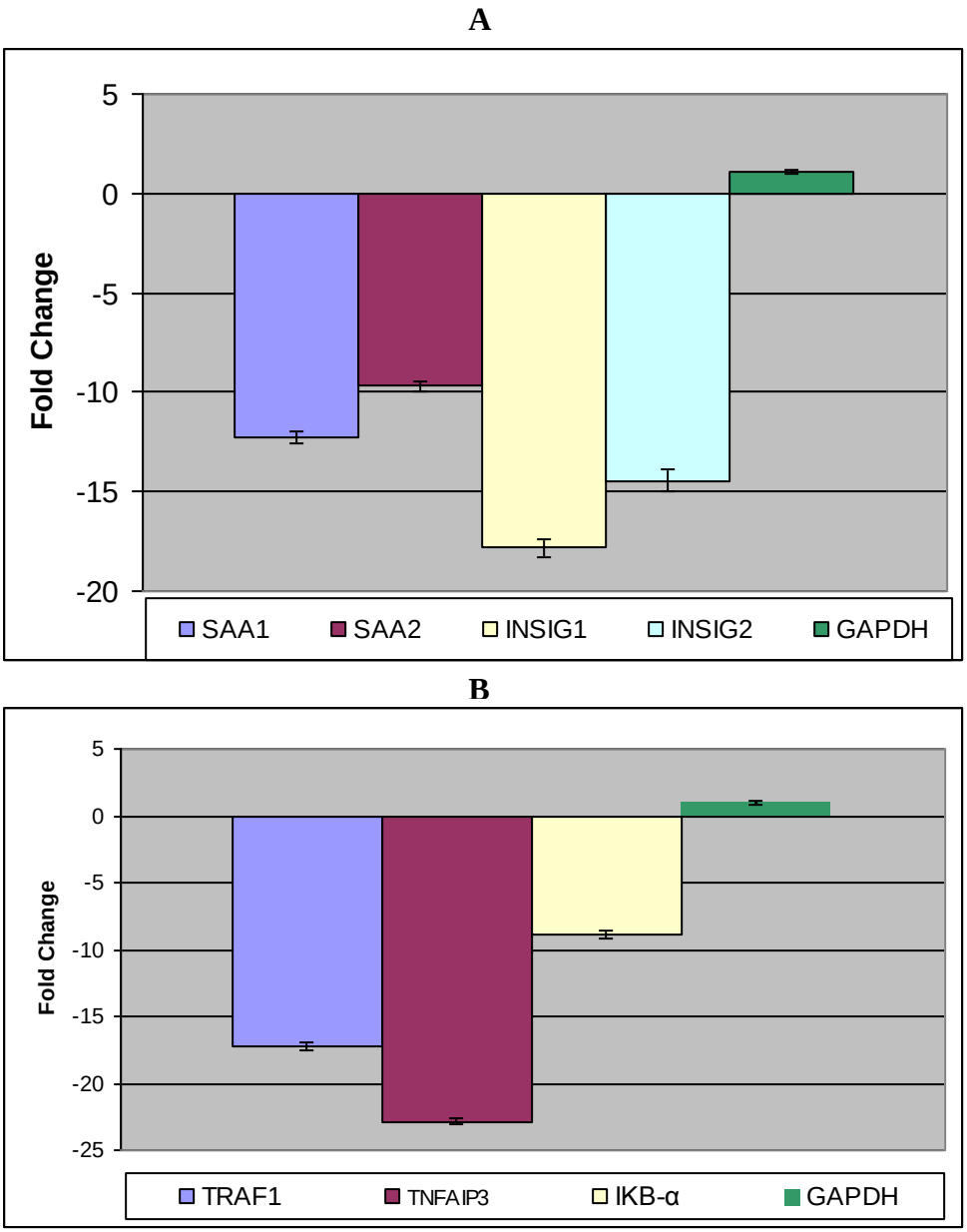


Figure F4.2. Quantitative PCR Analysis. mRNA levels of

A) sterol regulatory genes and
B) TNFR2 signaling pathway genes we assessed in 11.8 hybrid and 11.8_ts segregant cell lines by quantitative real-time PCR. The calculated fold change (See Appendix I 1.12) indicates fold repression in the segregant line.

4.1.3 Adaptive-quality based clustering of differentially expressed genes

Genes that share common biological endpoints are often precisely transcriptionally co-regulated. Thus, to gain further biological insights, we studied the expression patterns of the differentially expressed genes by cluster analysis. We applied a method of adaptive quality-based clustering to group genes according to the degree of similarity in their microarray expression profiles. Compared to other clustering algorithms, adaptive quality-based clustering has the advantage of grouping only the most significantly related genes without forcing genes with otherwise low correlations into common clusters [89]. Using a stringent probability of cluster membership of 0.95, we observed a highly co-expressed cluster of 20 genes down-regulated in parental (PvH) and segregant (HvS) cell lines that contained almost all the sterol metabolic genes identified in the previous GO and pathway analysis (Tables T4.6 A and T4.6 B). Interestingly, this gene cluster also contained the TNFR2 signaling pathway genes: A20 (TNFAIP3), IKB-α and TRAF1, indicating, via co-expression, a possible transcriptional link between TNFR2 signaling and sterol metabolism in the maintenance of the nontumorigenic phenotype.

GENE NAME	GENE SYMBOL	SRE*	P-Value^
START DOMAIN CONTAINING 4, STEROL REGULATED	STARD4	2	4.10E-04, 1.10E-04
STEROL-C5-DESATURASE (ERG3 DELTA-5-DESATURASE HOMOLOG, FUNGAL)-LIKE	SC5DL	1	2.90E-04
3-HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE	HMGCR	2	2.02E-03, 1.80E-04
7-DEHYDROCHOLESTEROL REDUCTASE	DHCR7	2	9.10E-04, 2.40E-04
MEVALONATE (DIPHOSPHO) DECARBOXYLASE	MVD	1	<1e-05
TNF RECEPTOR-ASSOCIATED FACTOR 1	TRAF1	1	1.00E-05
SQUALENE EPOXIDASE	SQLE	2	1.10E-04, 3.40E-04
FATTY ACID DESATURASE 2	FADS2	1	3.90E-04
FARNESYL-DIPHOSPHATE FARNESYLTRANSFERASE 1	FDFT1	2	3.00E-05, 3.90E-04
ISOPENTENYL-DIPHOSPHATE DELTA ISOMERASE 1	IDI1	2	6.00E-05, 1.90E-04
INSULIN INDUCED GENE 1	INSIG1	1	1.0E-05
TUMOR NECROSIS FACTOR, ALPHA-INDUCED PROTEIN 2	TNFAIP2	1	3.10E-04
SERINE PALMITOYLTRANSFERASE, LONG CHAIN BASE SUBUNIT 1	SPTLC1	2	<1e-05, 2.00E-05
STEROL-C4-METHYL OXIDASE-LIKE	SC4MOL	1	4.00E-05
PHOSPHOLIPASE A2 RECEPTOR 1, 180KDA	PLA2R1	1	4.50E-04
HISTONE DEACETYLASE 9	HDAC9	2	1.66E-03, 1.68E-03
HYDROXYSTEROID (17-BETA) DEHYDROGENASE 7	HSD17B7	2	1.00E-04, 3.30E-04
ACYL-COA SYNTHETASE SHORT-CHAIN FAMILY MEMBER 2	ACSS2	2	2.31E-03, 7.80E-04
NUCLEAR FACTOR OF KAPPA LIGHT POLYPEPTIDE GENE ENHANCER IN B-CELLS INHIBITOR, ALPHA	NFKBIA(IKB-α)	2	4.2E-04, 7.0E-05
TUMOR NECROSIS FACTOR, ALPHA-INDUCED PROTEIN 3	TNFAIP3	1	1.0E-05

*the number of Sterol Regulatory Element (SRE) motifs identified in promoter regions.
^P-value indicates the number of times a randomly generated pattern of the same length (as that of SRE) scores a PWM score better than the SRE motif.

Enriched GO Terms	No of Genes	% of genes in the cluster	Fisher's Exact P-Value
sterol biosynthesis	8	5.3	9.1E-10
steroid biosynthesis	10	6.6	6.1E-9
lipid biosynthesis	14	9.2	1.5E-8
cellular lipid metabolism	17	11.2	9.8E-7
sterol metabolism	8	5.3	1.4E-6
steroid metabolism	10	6.6	3.4E-6
lipid metabolism	18	11.8	4.9E-6
cholesterol biosynthesis	5	3.3	2.0E-5

Table T4.6: AQBC analysis results.

A) Members of a 20-gene cluster obtained through the Adaptive-quality based clustering (AQBC) algorithm.
B) Gene ontology enrichment analysis of the 20-gene cluster generated by AQBC.

4.1.4 In silico promoter analysis of co-regulated genes

That the expression patterns of genes comprising the AQBC cluster were highly correlated suggests the possibility of transcriptional co-regulation by a common transcription factor. Interestingly, several of the sterol metabolism genes identified in this cluster are known to be transcriptionally regulated by the Sterol Regulatory Element Binding Proteins (SREBP) [90]. SREBPs are basic helix-loop-helix leucine zipper transcription factors that bind to sterol regulatory elements (SREs) with DNA sequence "TCACNCCAC" located in the promoter regions of sterol metabolism genes[91-93].

Using a web-based DNA sequence motif location and discovery tool we developed at the Genome Institute of Singapore (termed BEARR1.0, and which is the subject of the next chapter), we analyzed the promoters of the genes within the AQBC 20-gene cluster for putative SREs. The transcription factor matrix for SREBP as defined in the TRANSFAC transcription factor database (http://www.genome.jp/dbget-bin/show_tfmatrix) was used as input for the BEARR 1.0 promoter analysis (Appendix I Section 1.16). Intriguingly, all 20 genes were found to possess at least one putative SRE (with 11 genes possessing two SREs), including the three TNFR2 pathway-related genes: TRAF1, TNFAIP3 and IKB-α.

4.1.5 Western blot and protein localization studies of the NPC cell lines

While SREBPs are known to regulate the expression of sterol metabolism genes, they have not been implicated in the regulation of genes of the TNFR2 pathway, prompting us to further explore this possibility. The SREBPs are comprised predominantly of two genes, SREBP1 and SREBP2, with similar biological functions. While our microarray did not contain probes for assessing SREBP1 expression, the SREBP2 probes indicated no detectable expression of SREBP2 mRNA. Thus, we focused our investigation on the potential involvement of SREBP1. First, we asked whether the SREBP1 protein was differentially expressed or showed different subcellular localization patterns between nontumorigenic hybrids and tumorigenic segregants. Physiologically, the SREBP1 protein exists in one of two biological states: a transcriptionally inactive, unmodified 128 kDa protein localized to the nuclear envelope and endoplasmic reticulum, or a transcriptionally active form as a cleaved 68 kDa polypeptide localized to the nucleus [94].

Western blot analysis of SREBP1 revealed markedly higher levels of the 68 kDa (active) peptide in hybrids (compared to segregants) and substantially higher levels of the 128 kDa protein (inactive) in segregants relative to hybrids (Figure F4.3A) (Appendix I 1.17). By immunofluorescent staining, we observed that SREBP1 is predominantly restricted to the nuclear envelope/ER in a segregant line (Figure F4.4A) and localizes to the nucleus in the corresponding hybrid line (Figure F4.4B). Together, these observations strongly suggest that SREBP1 is transcriptionally active in hybrids and transcriptionally inactive in segregants.

To assess the dependency of TRAF1, TNFAIP3 and IKB-α expression levels on SREBP1 activity, we conducted siRNA-mediated SREBP1 silencing assays on the hybrid cell lines. Knock down of SREBP1 in 11.8 hybrid cells resulted in a substantial decrease in the expression of the 68 kDa active form (Figure F4.3B), loss of nuclear staining (Figure 4.4C), and greater than a 25-fold reduction in the expression levels of TRAF1, TNFAIP3 and IKB-α (Figure F4.5) compared to the same hybrid cells transfected with a scramble siRNA (control). These results suggest that the expression levels of all three TNFR2 pathway-related genes are directly dependent on SREBP1 expression.

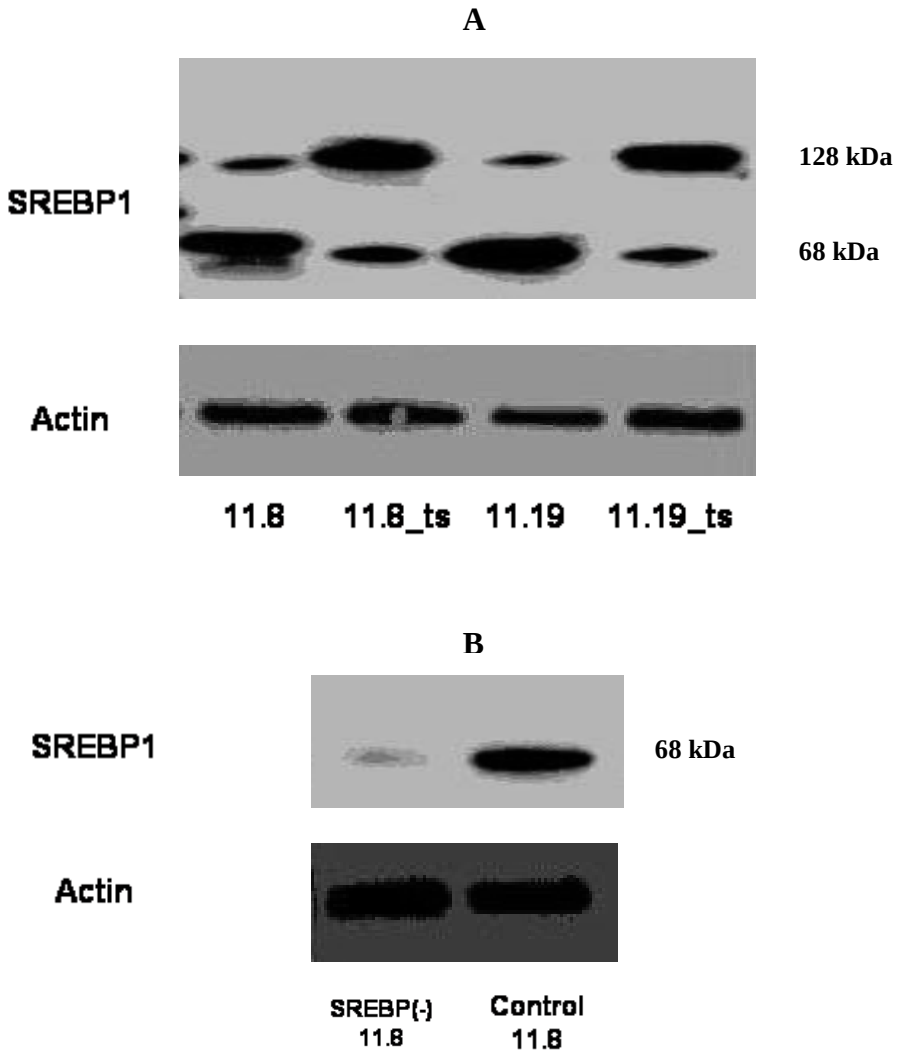
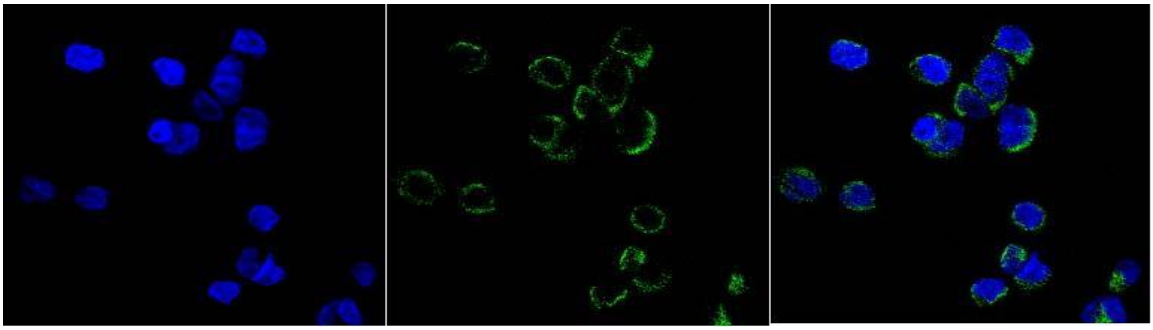


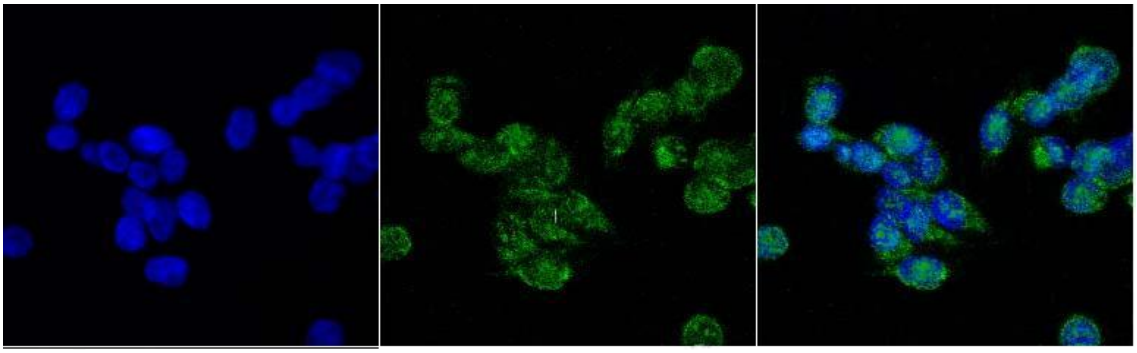
Figure F4.3: Western blot of SREBP1.

A) Protein levels of the inactive (128 kDa) and (68kDa) SREBP1 polypeptides in hybrid (11.8, 11.19) and corresponding segregant (11.8_ts, 11.19_ts) cell lines.
B) Levels of the active form of SREBP1 are shown in the 11.8 hybrid (siRNA scramble control)line and the 11.8 hybrid line with siRNA-mediated knock down of SREBP1 (SREBP(-)).

A



B



C

Figure F4.4: Immunofluorescent staining of SREBP1. Shown is SREBP1 staining in

A) 11.8_ts segregant and
B) 11.8 hybrid cell lines.
C) Shown is absence of SREBP1 staining in 11.8 hybrid cells silenced for SREBP1 expression by siRNA. DAPI (nuclear) stain is shown in blue, SREBP1 is stained in green.

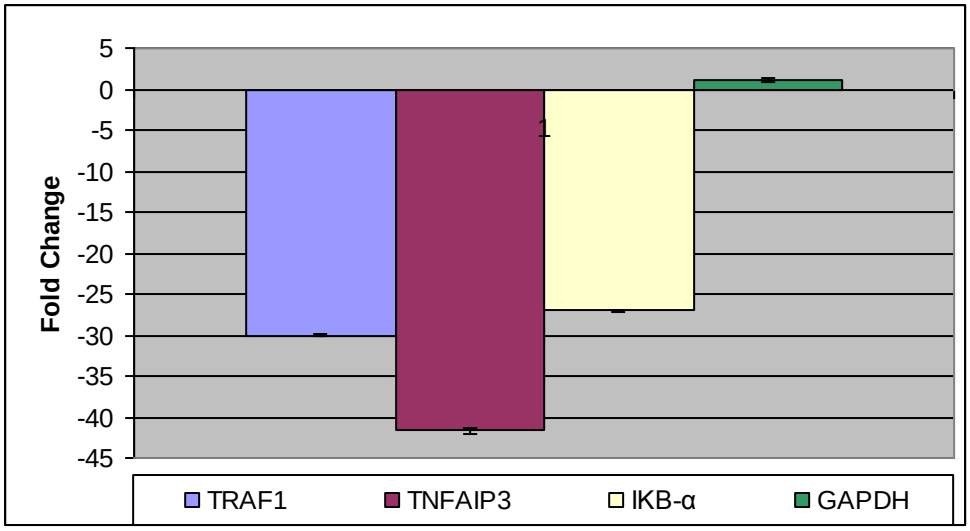


Figure F4.5: Quantitative expression analysis of TNFR2 signaling pathway genes.

11.8 hybrid cells transfected with a scramble (control) siRNA or depleted of SREBP1 by siRNA-mediated knock down were assessed for TRAF1, TNFAIP3 and IKB-α expression by quantitative real-time PCR. The calculated fold change indicates fold repression associated with SREBP1 knock down.

4.1.6 Testing for binding of SREBP1 to gene promoters

To test for a physical interaction between SREBP1 and the promoters of TNFR2 pathway genes bearing putative SREs, ChIP assays were carried out on the 11.8 and 11.19 hybrids (see Appendix I 1.22). Briefly, immunoprecipitates from SREBP1 and mock pull downs were compared for genomic DNA (promoter) detection by RT-PCR using primers specific for DHCR7, TNFAIP3, TRAF-1 and IKB-α. DHCR7 is a known SREBP-regulated gene and was used as a positive control. Because the IKB-α promoter harbors two putative SRE binding sites, RT-PCR was carried out on each SRE separately. In both hybrid lines, we observed a positive enrichment for the promoter regions of all the genes in the SREBP1 pull down, suggesting that SREBP1 binds to the promoters of each of the genes tested (Figure F4.6A & B). The extent of binding, however, appeared variable, with the strongest interaction involving TNFAIP3.

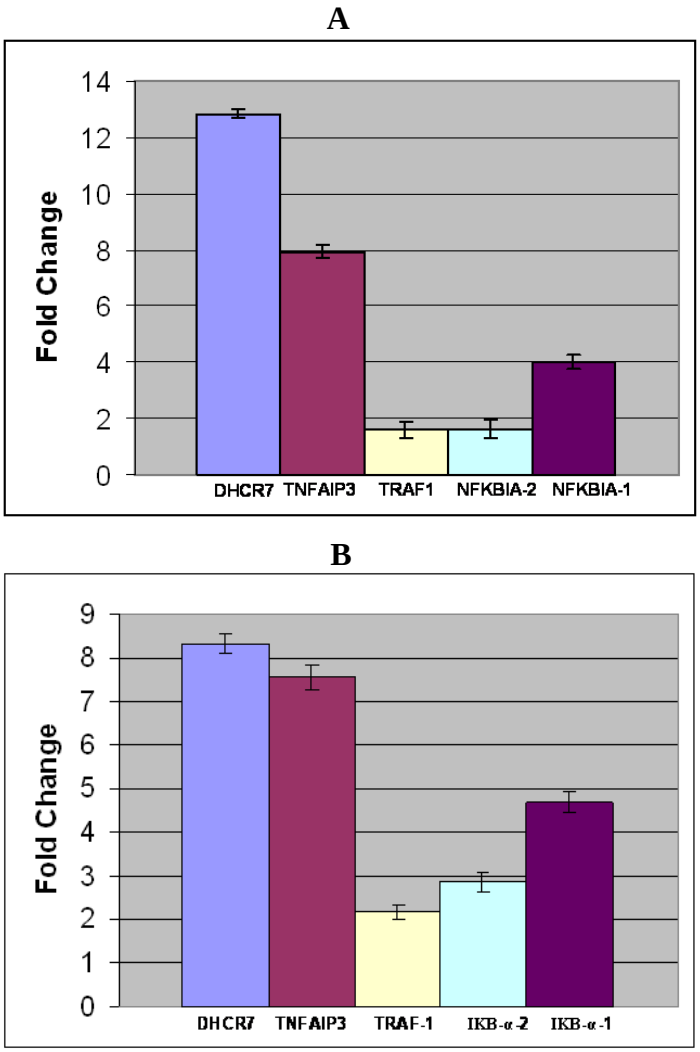


Figure F4.6: Detecting protein-DNA interactions by SREBP1 CHIP and RT PCR.

Immunoprecipitates from SREBP1 and mock pull downs in
A) 11.8 and
B) 11.19 hybrids were compared by RT PCR for the expression of select genes. Positive fold change indicates enrichment in the anti-SREBP1 samples.

4.1.7 SRE-dependent transcription of TNFR2 pathway genes

For further evidence that SREBP1 transcriptional activity is responsible for the observed expression of TNFR2 pathway genes, we carried out luciferase assays using the regulatory regions of TRAF1, TNFAIP3 and IKB-α containing the observed wild type SRE motif or a *mutant SRE* altered by site-directed mutagenesis (See Appendix I 1.20). In a comparison of 11.19 (hybrid) and 11.19_ts (segregant) cells, the wild type promoters of TNFAIP3 and TRAF1 were significantly transcriptionally repressed in the segregant cells (Figure F4.7A), while repression of the TNFAIP3 and IKB-α wild type promoters was observed in the 11.8_ts segregant (compared to the 11.8 hybrid) (Figure F4.7B). While somewhat variable, these results were mostly consistent with the differential expression observed of the endogenous genes. Next, we compared the luciferase activity of constructs bearing wild type and mutated SREs in the hybrid lines. Expression off the TNFAIP3 promoter was highly sensitive to SRE mutational status, and showed a highly significant reduction in activity from the mutant SRE in both 11.19 and 11.8 hybrids (Figure F4.7C & D). The TRAF1 and IKB-α promoters also showed a reduction in activity in the presence of the mutant SRE in the 11.8 hybrid, although to a lesser extent than TNFAIP3 (Figure F4.7D). Notably, promoter analysis of IKB-α identified two putative SRE sites, and a double-site disruption was required to effect a prominent transcriptional repression in the hybrids (Figure F4.7C & D inset). Together, these results suggest that TNFAIP3 and IKB-α promoters (and perhaps to a lesser extent TRAF1) are dependent on the wild type SRE sequence for transcriptional activation.

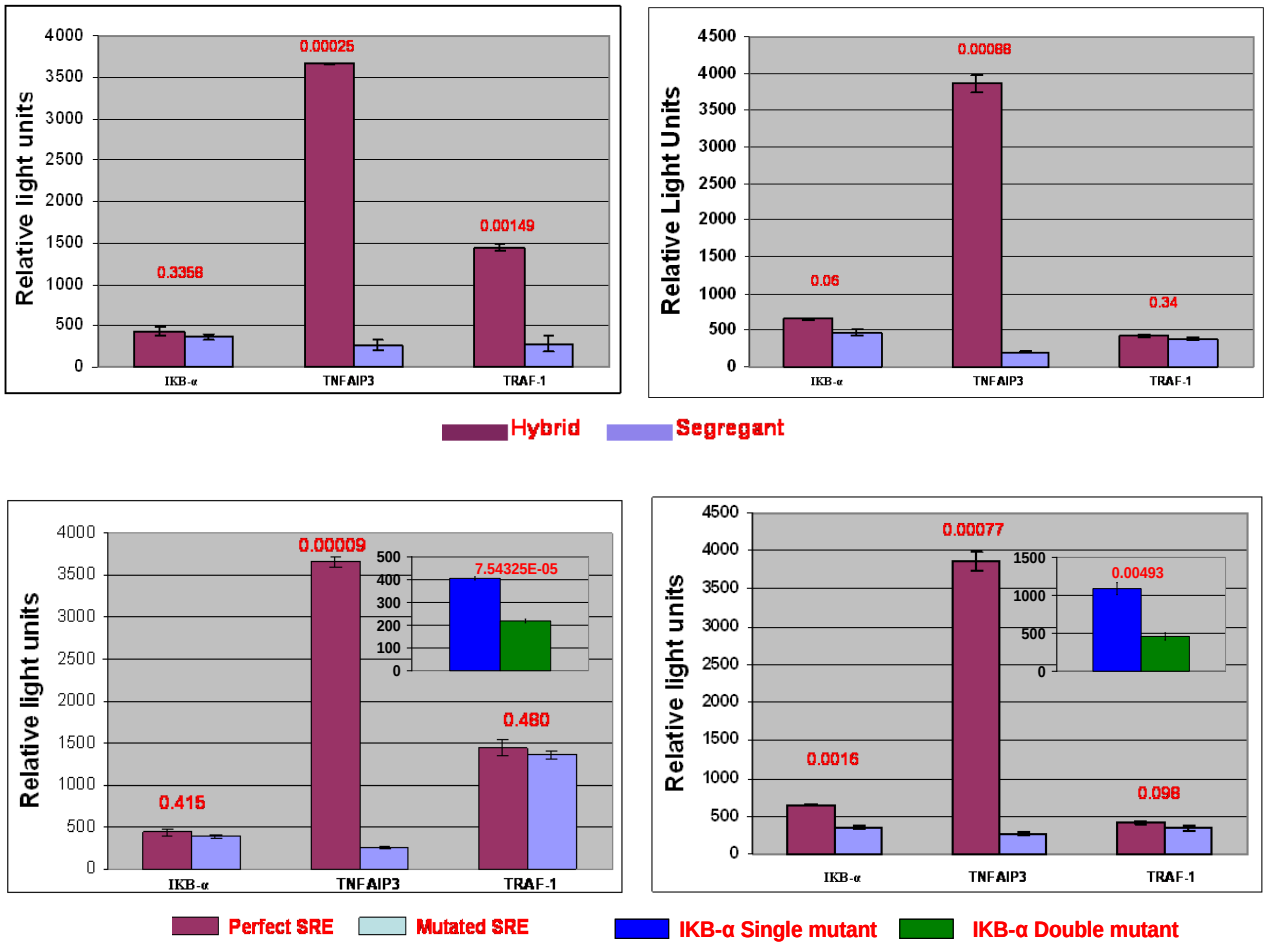


Figure F4.7: Transcriptional activation from SRE wild type and mutant promoters.

Luciferase assays were conducted using promoter regions of TRAF1, TNFAIP3 and IKB-α with or without mutant SREs.
A) Comparison of luciferase activity between hybrid and segregant (11.19) cells transfected with wild type SRE constructs.
B) Comparison of luciferase activity between hybrid and segregant (11.8) cells transfected with wild type SRE constructs.
C) Comparison of luciferase activity between hybrid cells (11.19) transfected with wild type or mutant SRE constructs.
D) Comparison of luciferase activity between hybrid cells (11.8) transfected with wild type or mutant SRE constructs. The inset in **C & D** shows the result of transfecting with a IKB-α double-mutant construct harboring mutations in each of its two putative SREs. P-values shown in red indicate the statistical significance of the difference between the two samples.

4.2 Discussion

In our MMCT mouse xenograft model of nasopharyngeal carcinoma, tumorigenic HONE1 cells revert to a nontumorigenic "hybrid" phenotype upon introduction of an exogenous chromosome 11, and subsequently, convert back to a tumorigenic "segregant" phenotype upon loss of critical regions of chromosome 11 and, as we have recently shown, the gain of other cytogenetic alterations. Concomitant with these genomic changes, we have observed hundreds of reproducible transcriptional alterations that by GO analysis are indicative, significantly, of changes in discrete biological pathways including the sterol metabolism and TNFR2 signaling pathways.

AQBC cluster analysis of the differentially expressed genes revealed a high correlation between the expression patterns of the sterol metabolism and TNFR2 pathway genes, suggesting the possible transcriptional co-regulation of these otherwise unrelated signaling compartments. Sequence analysis of the gene promoters, provided further support of this hypothesis, as it revealed the existence of one or more sterol regulatory elements (SREs), the binding motif recognized by the sterol-regulating SREBP transcription factors, in all (20/20) of the clustered genes. Thus, we explored the possibility that an SREBP transcription factor might not only modulate the expression of sterol metabolism genes [90], but also TNFR2 pathway genes, namely, TRAF1, TNFAIP3 and IKB-α, in a manner associated with NPC tumorigenicity.

First, we observed evidence by Western blot analysis and immunofluorescent staining that SREBP1 is transcriptionally active in hybrids and transcriptionally inactive in segregants. In hybrids, SREBP1 existed predominantly in the 68 kDa *active* form localized to the nucleus,

while in segregants; SREBP1 was observed localized to cytoplasmic perinuclear structures in the *inactive* 128 kDa form. Next, we observed that siRNA-mediated knock down of SREBP1 in hybrid cells resulted in a dramatic loss of TRAF1, TNFAIP3 and IKB-α expression. We then observed evidence by ChIP-RT PCR that SREBP1 can physically interact with and "pull down" the promoter regions of all three genes. Finally, we created luciferase constructs using the SRE-containing promoter regions of these genes, and demonstrated their transcriptional sensitivity to the wild type SRE sequence. In most cases, transcription was repressed in the presence of a mutant SRE. Together, these data are consistent with the hypothesis that TRAF1, TNFAIP3 and IKB-α, key signaling molecules in the TNFR2 pathway, are transcriptionally activated by SREBP1 in nontumorigenic hybrids, and transcriptionally repressed through loss of SREBP1 activation in tumorigenic segregants.

TNFR2 is the receptor for Tumor Necrosis Factor TNF (a.k.a. Lymphotoxin), a cytokine normally produced by activated lymphocytes. TNF mediates a large variety of inflammatory, immunostimulatory, and antiviral responses, and is cytotoxic to many tumor cell types. TNF receptor (TNFR) associated factors, such as TRAF1, associate with and mediate the signal transduction from TNFR2 and other receptors of the TNFR superfamily. In complex with TRAF2 and TNFRs, TRAF1 is required for TNF-mediated activation of MAPK8/JNK and NF-kappa B. NF-kappa B is a transcription factor complex that plays a key role in modulating cell survival signals. Downstream of TNF/TRAF signaling, activated NF-kappa B is, paradoxically, known to protect cells from apoptosis induced by TNF-alpha, ionizing radiation and daunorubricin [95] rather than potentiate cell death. This protective effect is achieved, in part, through NF-kappa B-mediated suppression of caspase-8 activation [96, 97]. More recently, NF-

kappa B has been implicated in the potentiation of angiogenesis. Fujioka and colleagues [98] observed that NF-kappa B is constitutively activated in pancreatic adenocarcinoma, and that inhibition of NF-kappa B by IKB-α suppressed metastatic progression of AsPc-1 cells (pancreatic cancer) in nude mouse xenografts and decreased neoplastic angiogenesis in tumors via reduced expression of Vascular Endothelial Growth Factor (VEGF). NF-kappa B-mediated induction of angiogenesis through VEGF expression has since been observed in several cancer types [99, 100]. Together, these data indicate that activated NF-kappa B can promote tumor growth and progression through promoting both anti-apoptotic and pro-angiogenic signals.

Of the TNFR2 pathway genes we discovered to be transcriptionally activated by SREBP1 in the nontumorigenic hybrids, increased expression of TRAF1, a known activator of NF-kappa B, would appear to be more a promoter of cancer progression rather than tumor suppression. However, both TNFAIP3 and IKB-α, also transcriptionally activated by SREBP1 in the nontumorigenic hybrids, have well defined roles as potent suppressors of NF-kappa B activity [101, 102]. TNFAIP3 is thought to inhibit NF-kappa B through ubiquitin-mediated proteolysis [103], while IKB-α binds directly to NF-kappa B silencing its transcriptional activity through cytoplasmic sequestering [104]. Importantly, both proteins have been shown to suppress tumorigenic properties of cancer cells. Zhang et al [105] demonstrated that TNFAIP3 expression inhibits invasion of a salivary carcinoma cell line through inactivation of NF-kappa B, while antisense RNA suppression of IKB-α was shown to induce malignant transformation of NIH3T3 cells [106].

Our data, though preliminary, suggest a model of tumor suppression whereby the transcriptional activity of SREBP1, the primary regulator of sterol biosynthesis, suppresses activation of NF-kappa B through induction of TNFAIP3 and IKB-α. The question thus arises, how then might loss of SREBP1 activity occur in NPC, leading to the tumorigenic phenotypes of the parental and segregant cells?

Cholesterol is a ubiquitous lipid sterol that plays a key role in the composition of cell membranes and the synthesis of steroid hormones. As excess cholesterol can have cytotoxic effects, the precise monitoring of its intracellular levels is of high physiological importance (Figure F4.8). When cholesterol reaches high intracellular levels, SREBPs are rapidly converted to the inactive form, shutting down the expression of cholesterol metabolism genes. Concomitantly, excess cholesterol is actively transported out of cells by the serum amyloid proteins SAA1 and SAA2. At low cholesterol levels, SREBPs are then shifted to the active form, thus promoting expression of SREBP target genes and further cholesterol production.

Cholesterol Biosynthesis

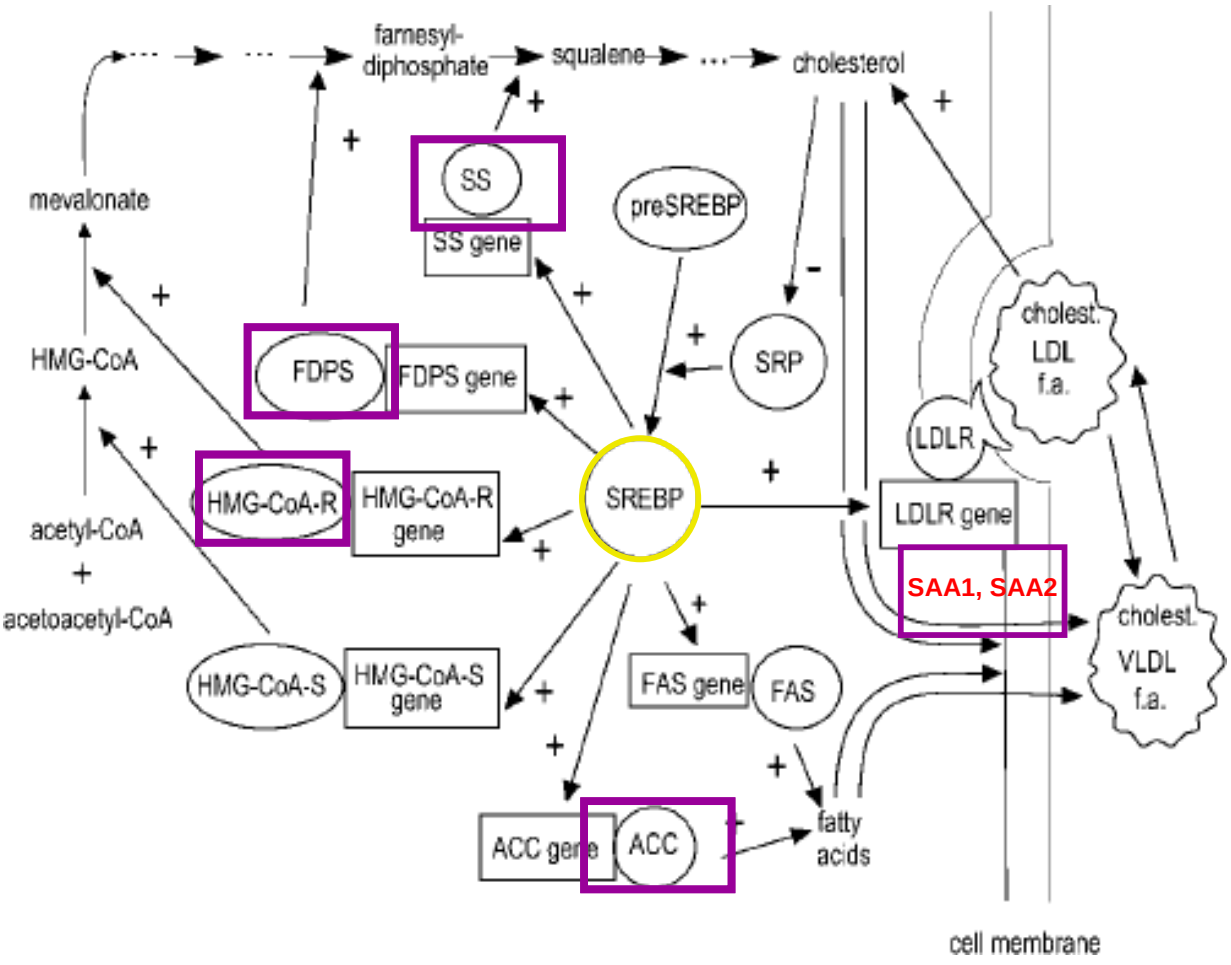


Figure F4.8: Diagram showing the cholesterol biosynthesis pathway.

Genes that were down regulated in the cancer phenotype parental and segregant cell lines are enclosed in purple rectangles

Interestingly, the key regulators of cholesterol levels, SAA1 and SAA2, are located together at the 11p15.1 locus (Figure F4.1) which is absent in the parental cells but present in the hybrids (by virtue of MMCT). By density plot analysis, we observed evidence of a prominent loss of this same region in the tumorigenic segregant lines (Chapter 3). Taken together, these observations are consistent with the hypothesis that loss of SAA1 and SAA2, due to 11p15.1 deletion in the parental and segregant cell lines, gives rise to high intracellular cholesterol levels that repress SREBP1 transcriptional activity. This, in turn, prevents SREBP1-mediated expression of TNFAIP3 and IKB-α leading to the activation (via de-repression) of NF-kappa B in the parental and segregant lines contributing to their tumorigenicity.

To test this hypothesis, future research will focus on the transcriptional activity of NF-kappa B in our model, the tumor suppressive roles of SAA1 and SAA2 (can they rescue the nontumorigenic phenotype in parental and segregant cells?), and the impact of cholesterol reducing agents such as bile acid sequesterants, fibric acid derivatives and HMG-CoA inhibitors (i.e., "statins") in reverting the tumorigenic phenotypes through inactivation of NF-kappa B. Interestingly, statins are routine cholesterol-reducing drugs that have various pro-apoptotic and anti-metastatic activities that could play a role in cancer risk or progression. Prostate cancer growth has recently been linked with high cholesterol levels, and statins have been shown to have a curative effect on some prostate cancers [107, 108] as well as to significantly reduce the risk of advanced prostate cancer [109]. The intersect between SREBP1 activity and de-repression of NF-kappa B as potential downstream effectors of cholesterol-associated tumorigenicity warrants further investigation.

Chapter 5: Tools for Exploring the Vocabulary of Transcription Factors

5.0 Transcriptional Response Elements and Gene Regulation

Gene transcription is a prerequisite of all biological processes, with the essential function of producing RNA molecules, such as messenger RNAs that can be translated into the protein machinery of a cell. Regulation of transcription is controlled by transcription factors that recognize transcriptional response elements or "binding motifs" encoded in the nucleic acid sequence of gene promoters. Transcription factors are master regulators of transcriptional networks that dictate which biological processes are operational, or not, within a cell. Disruption of transcriptional networks is frequently observed in diseases such as cancer, and the study of these networks and how they are perturbed in different systems can help to unravel important components of biological and pathological processes.

Various genomic strategies such as expression microarrays, for example, enable the genome-wide comprehensive monitoring of transcriptional changes within cells. However, how these changes relate to distinct (if overlapping) transcriptional networks is not always intuitive. The mapping of specific transcription factor binding sites in the cis-regulatory regions of differentially expressed genes may allow the identification of direct targets of transcription factors, facilitate the parsing of genes into functional transcriptional networks, and identify mechanistically-relevant targets for downstream experimental and clinical manipulation.

5.1 Batch Extraction and Analysis of cis-Regulatory Regions (BEARR)

Some web-accessible computational tools for extracting genomic sequence data and analyzing transcription factor binding sites are currently available to researchers, although with varying degrees of accessibility, capability and focus. The most comprehensive tools, including TOUCAN [110], EZRetrieve [111], the Genomatix suite (www.genomtix.de) and TRANSPLORER (www.biobase.de) from BIOBASE (the former two are from academic sources and the latter two are only available commercially for batch analysis), usually employ annotation-based sequence retrieval and position weight matrices, typically from the TRANSFAC database[112] or specialized databases, for sequence extraction (e.g. the upstream region extraction tool developed by the Harvard-Lipper Center, http://arep.med.harvard.edu/labgc/adnan/hsmmupstream) and binding site identification, respectively. There are also stand-alone [e.g. MEME [113], BioProspector [114], MDScan and REDUCE] and integrated [e.g. MotifSampler and phylogenetic footprinting modules in TOUCAN] motif discovery tools designed to uncover conserved sequences in extracted regulatory regions [115, 116]. While the sophisticated approaches in these programs represent innovations in comprehensive binding site prediction and discovery and are useful for detailed analysis and mining of refined or idealized datasets—for example members of the same gene family or known targets of specific transcription factors—nevertheless, based on our experience, they may be of limited use against the large amount of 'noisy' data generated in microarray studies that often confound the interpretation of the analytical results or fail to yield conserved regulatory sequences. Whereas these tools favor comprehensive approaches for identifying all known binding sites or conserved sequence discovery in the dataset, we propose that a critical and complementary step prior to a more comprehensive analysis and experimental validation is a

hypothesis-driven and focused computational strategy based on the experimental design. Furthermore, although batch options are available for some of the tools noted, they have not been optimized for batch operations and are not suitable for the large amount of data that can be generated from genome-scale studies. Therefore, we created Batch Extraction and Analysis of cis-Regulatory Regions, or BEARR, to assist biologists in performing batch extraction and analysis of cis-Regulatory regions of hundreds or even thousands of differentially expressed genes identified in microarray studies. Here, we describe the system design and functionalities.

5.2 System design

The system is divided into two parts: (i) the regulatory region sequence extraction module and (ii) the sequence analysis module. Essentially, the sequence extraction module generates the nucleotide sequences, using annotation and genomic sequence databases, to be queried by the analysis module, based on transcription factor binding sites or any conserved regions defined by the user. The modular system design allows for maximum component reusability and extensibility. To ensure platform independence and ease of installation, no specialized library or programs were employed (Figure F5.1A & F5.1B). An interactive web-based user interface was developed for ease of use, interoperability and accessibility (Figure F5.2).

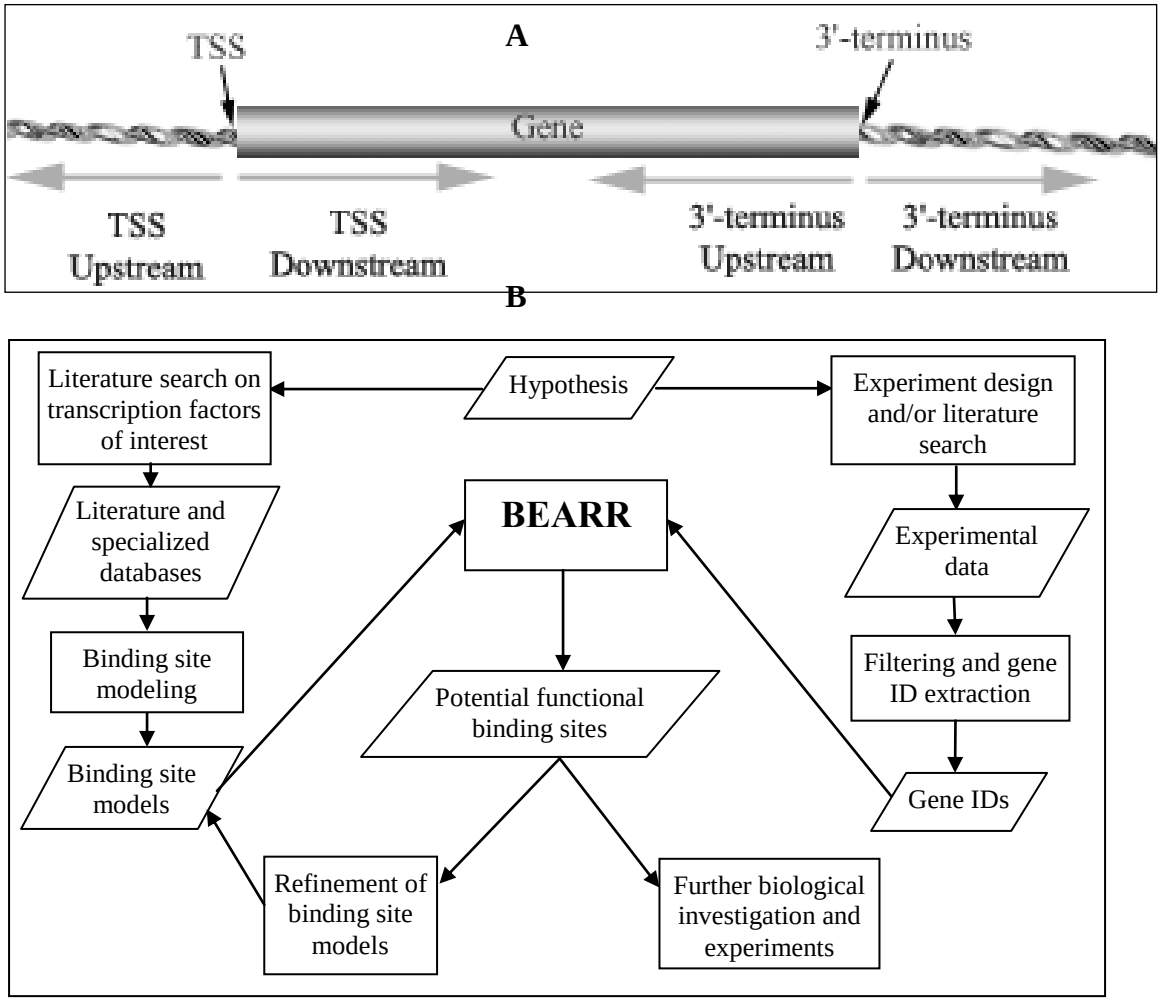


Figure F5.1: BEARR 1.0 workflow.

A) Cartoon representation of the genes upstream and downstream regulatory regions that can be extracted using BEARR 1.0 tool.
B) Schematic representation of the work flow that could be adopted using BEARR 1.0 tool.

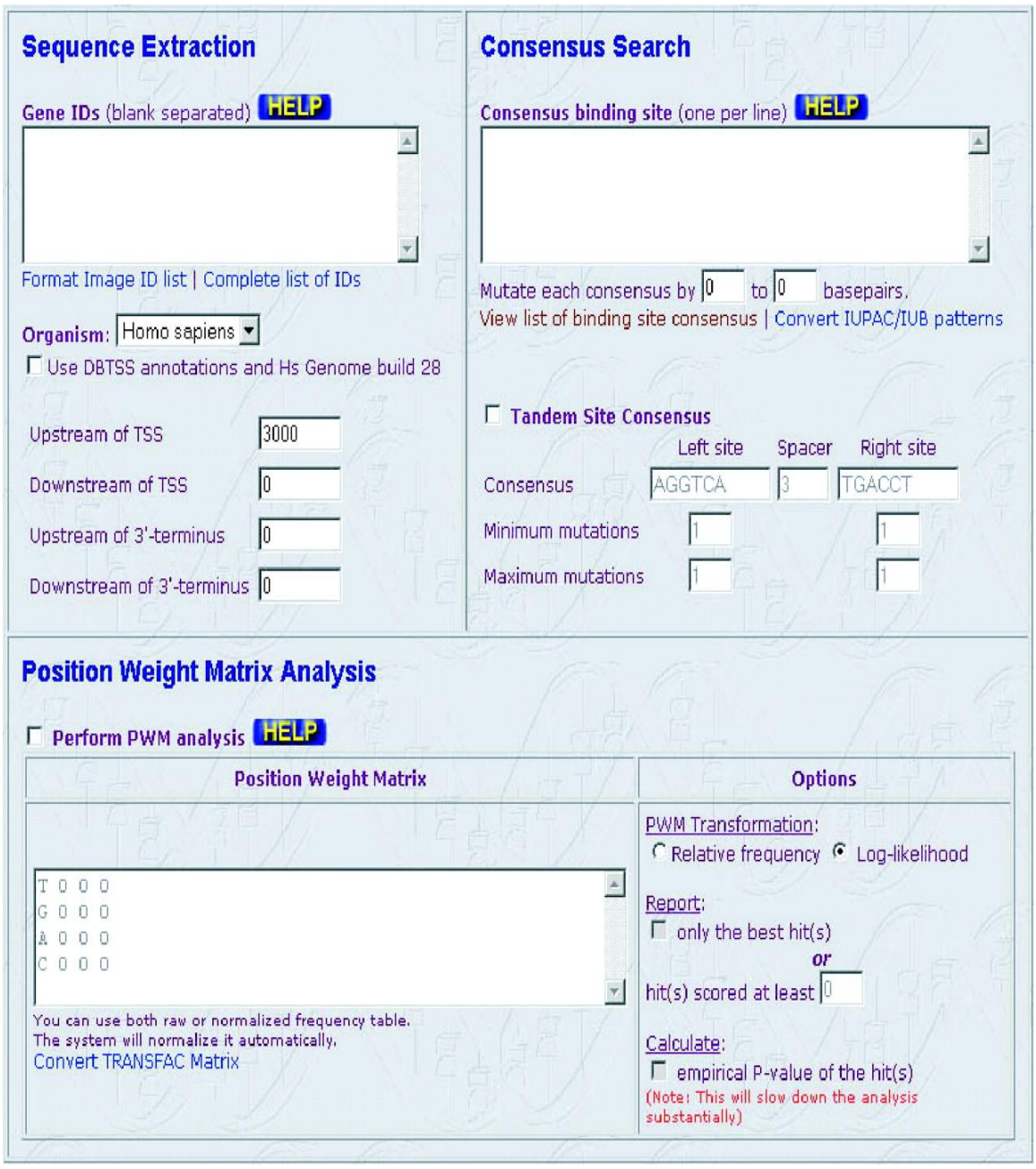


Figure F5.2: Screen shot of the graphical interface and the components of the page

5.3 Regulatory region extraction

RefSeq accession numbers, recognizable by the NM_ and XM_prefixes, are the primary gene identifiers employed in the modules, although the system is capable of translating other identifiers from microarray analysis tools, including the UniGene cluster ID and the IMAGE clone ID. The list of gene identifiers can be easily copied from the output files of most expression analysis tools and pasted into the input window of the extraction module. This input feature allows users the flexibility of utilizing the optimal microarray data analysis tools to identify differentially expressed genes and then efficiently moving into the extraction and analysis of their respective regulatory regions. Currently, BEARR accepts identifiers for human and mouse genes.

The sequence extraction module utilizes NCBI's LocusLink and RefSeq [117] databases to identify the genes and pinpoint their loci within the genome. These annotations were chosen for their comprehensiveness, in terms of number of annotated genes, and their consistency with the current state of the NCBI contig databases, the underlying genomic sequence database used in BEARR. Using the transcription start site (TSS), defined as the 5'-most nucleotide in the reference transcript, and the 3' terminus of the transcript as reference points, the module can extract user-defined regions both upstream and downstream of the references. Extraction speed has been enhanced by utilizing downloaded databases on the BEARR server rather than accessing the information on the NCBI servers via the internet. The resulting sequences are saved in the FASTA format and are downloadable by users. It is important to note that the TSS locations annotated in LocusLink and RefSeq may only approximate true start sites due to incomplete information at the 5' ends of some reference sequences. The large number of

annotated genes in LocusLink and RefSeq, however, provides the advantage of maximizing the number of regulatory regions we are able to extract for further analysis. To better annotate human start sites, we have incorporated the information from DBTSS [118], a database of experimentally extended 5' sequences in human transcripts, to assist users in orienting the locations of predicted sites with reference to the more accurate TSS. It is also expected that as the sequence and annotation information is updated regularly in LocusLink and RefSeq, the accuracy of the current extraction method will improve.

5.4 Sequence analysis

Once the desired sequences are obtained by the extraction module, the sequence analysis subsystem queries the input for putative transcription factor binding sites. Although it may be possible to map all known binding sites by incorporating information from TFD [119], TRANSFAC [112] and TRRD [120], BEARR is specifically designed to efficiently detect a focused group, as determined by experimental design or the microarray data, of transcription factor binding sites. We have, however, provided links to the information in TFD and TRANSFAC to enable users to extract consensus sites or matrices of their choice for use in BEARR. Users also have the option of inputting their own binding sites derived from the latest experimental data rather than relying on the potentially dated information in the databases. Extracted sequences can be queried using single or multiple consensus binding sites, including tandem sites, or position weight matrices of binding sites [PWMs; [121]. Stringency is adjusted by defining the number of mismatches or spacer sequences allowed in the consensus searches or the similarity score threshold in the PWM output. The PWM score signifies the goodness of the match, where higher scores correspond to better matches. An optional empirical P-value

calculation is also available for users who wish to assess the statistical significance of the motifs being identified against the null hypothesis of finding the motifs in randomly generated sequences. The analysis results are displayed in a tab-delimited text file and include the detected putative binding sites, number of hits and their positions relative to either 5' end or 3' terminus. Additional columns of PWM score, empirical P-value, and DBTSS 5'-extension are also included under the appropriate settings (Figure F5.3)). Further information of the analysis module can be found in the Help sections of the website.

	A	B	C	D	E	F	G	H	I	J
1	Gene Name	Region	Strand	DBTSS 5'-extens	Accession numb	Pattern	PWM score	Empirical p-valu	Number of	Positions
2	AGT	5'-end	sense	378	NM_000029	AAGTGGCTGGGGCAA	1.57618189	0.00492	1	-2929 to -2915
3	AGT	5'-end	sense	378	NM_000029	GGGGCAAAGCTACAA	0.4775696	0.0075	1	-2921 to -2907
4	AGT	5'-end	sense	378	NM_000029	GGCTTATTTGGTTTC	1.18362019	0.00574	1	-2333 to -2319
5	AGT	5'-end	sense	378	NM_000029	CAGTCAGTTTGAATG	3.20957304	0.00235	1	-2146 to -2132
6	AGT	5'-end	sense	378	NM_000029	TAACCAAAATTATTT	0.19833658	0.00841	1	-1945 to -1931
7	AGT	5'-end	sense	378	NM_000029	TGAATGGAGTGATCA	0.32341892	0.00797	1	-1787 to -1773

Figure F5.3: Sample output from BEARR 1.0.

Sample output from BEARR when used to identify potential functional Estrogen Response Element, based on TRANSFAC ERE position weight matrix. Output of consensus pattern searches is similar, with the exception of PWM score and empirical P-value columns.

In some cases, the user may find it necessary to redefine the published consensus sequences or matrices for binding sites because the reported derivations may be based on outdated or incomplete information. It is also highly recommended that the users optimize the stringency thresholds, using a test set of known target genes of the transcription factors of interest, prior to the analysis of the extracted sequences. These parameters will directly impact the sensitivity and the specificity of the analysis. For further detailed and comprehensive analysis of the extracted sequences, users can upload the FASTA sequence files into other available regulatory region analysis tools [e.g. MEME] described above from the links on the website.

5.5 Discussion

BEARR was originally created to assist the discovery of potential estrogen receptor binding sites in the regulatory regions of hormone-responsive genes identified in our microarray experiments. We have subsequently used BEARR to analyze expression data from studies of other nuclear receptors and transcription factor families in order to identify target genes and downstream pathways. In general, this software is well suited for the analysis of data generated from study designs where transcription factor activity is manipulated experimentally. The software has also been applied in the analysis of the role of transcription factors identified in microarray studies of clinical samples in mediating the observed alterations in gene expression profiles of diseased tissues. In addition to the analysis of regulatory regions of genes identified in array studies, genome-wide surveys (all genes annotated in LocusLink and RefSeq) of transcription factor binding sites of interest, within defined regulatory regions, have been carried out to determine the frequency and distribution of putative binding sites. In a typical analysis flow, a user starts by formulating the underlying hypothesis, followed by designing and carrying

out the appropriate experiments while at the same time studying the literature for related transcription factor binding sites to construct an initial binding site model. The experiments enable the user to group interesting genes, from which BEARR tries to identify potential functional binding sites based on the input transcription factor binding site model. Refinement of the model might also be necessary. Further biological investigations and experimentations could be performed on the binding sites found. Examples, tutorials and analysis workflows can be found in the BEARR web page. As more genomes are sequenced and annotated in LocusLink and RefSeq, additional databases will be integrated into BEARR to enable users to extract and analyze regulatory regions of genes from other model organisms. These upgrades will enhance the comparative analysis of conserved transcription factor binding sites.

Chapter 6: Concluding Remarks

Cancer, in its various forms, is the leading cause of death in men and women worldwide; yet, the molecular mechanisms of cancer formation and progression remain poorly understood. While traditional methods in molecular biology and cytogenetics have led to the discovery of some important and therapeutically targetable cancer genes including Her2/neu (for breast cancer) and Gleevec (for lymphomas) [122], the discovery process has, in recent years, been greatly expedited by the use of cutting edge genomics and proteomics technologies, coupled with the recent elucidation of the human genome sequence. For example, through intelligent interrogation of microarray expression data, many genes with previously unknown links to tumorigenesis have been identified as having important roles in cancer prognosis, diagnosis and treatment [123].

In this work, we have combined experimental and informatics-driven techniques to devise a novel approach to the identification of cancer-associated genes. This approach combined the power of Microcell Mediated Chromosome Transfer (MMCT) and *in vivo* models of tumorigenesis, with the modern expression microarray and custom bioinformatics tools to discover, from a genomic perspective, genes and pathways with putative roles in nasopharyngeal carcinoma (NPC). From our analyses, we successfully identified and published the discovery of a putative tumor suppressor gene called THY-1 [124]. Through further investigation involving a biologically-guided data mining strategy, we discovered evidence of a novel link between cholesterol metabolism, SREBP1 activity and NPC, possibly involving the de-repression of NF-kappa B activation in NPC tumorigenesis.

For quite some time a connection between diet and cancer has been known, but the molecular details of this association have remained unclear. In the last decade a few studies have reported links between cholesterol, lipid levels and cancer, leading to investigations into the utility of cholesterol lowering drugs, such as the statins, in cancer treatment and prevention [107, 108]. Our observation of the possible involvement of SREBP1 inactivation, silencing of TNFAIP3 and IKB-α, and subsequent de-repression of NF-kappa B in NPC progression may shed light on the mechanisms underlying the cholesterol-cancer link, and warrants further investigation.

In conclusion, we posit that the intelligent analysis and integration of genomic read-outs of cancer behavior can drive the discovery of meaningful, and sometimes *profound*, molecular mechanisms with therapeutic potential. Beyond intelligent analytical design, perhaps the greatest challenge moving forward will be in the area of computational development – finding solutions for processing exponentially larger datasets and integrating the numerous platforms now available for measuring different aspects of genome dynamics. The post-genomic era is indeed perhaps only the beginning of a new age of computational systems-driven science that will have a lasting impact on the future landscape of *bio*-research.

APPENDIX I

1.0 NPC cell lines and culture conditions

The NPC HONE1 cells [125] used as the recipient cell line for monochromosome fusion were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% calf serum (DMEM/10% CS (5% Fetal CS + 5% Newborn CS); Life Technologies, NY, USA). The donor mouse hybrid cells MCH 556.15, containing an intact human chromosome 11 tagged with a selectable marker, neomycin resistance were grown in DMEM/10% CS containing 600 or 800 µg/ml of G418 (Geneticin) (Invitrogen, US) . Micro cell hybrid cell lines 11.8, 11.12, 11.13 and 11.19 derived by the fusion of HONE1 and MCH556.15 were selected in growth medium DMEM/10% CS containing 100 µg/ml streptomycin, 100 U/ml penicillin (Invitrogen, US) and 400 µg/ml G418. Tumor segregant cell lines 11.8-3TS, 11.12-2TS, 11.13-1TS, and 11.19-4TS which has been derived from their respective Hybrid lines were cultured in DMEM/10% CS containing 100 µg/ml streptomycin and100 U/ml penicillin (Invitrogen, US). All the cell lines are maintained at 37ºC in a humidified incubator under an atmosphere of 5% CO2 and 95% air.

1.1 Gene expression analysis using oligonucleotide microarray hybridization

A human oligo-library containing synthesized 60-mers of 18,912 genes was purchased from Sigma-Genosys (Woodlands, TX, US). These oligonucleotides were spotted on glass slides using GeneMachines Microarray spotter at the Genome Institute of Singapore.

To reduce variations in gene expression caused by culture conditions, we harvested total RNA from cells using the RNeasy Midi Kit (Qiagen, Hilden, Germany) at 70-80% confluency and 48 h after seeding. We produced cDNA labeled with Cy3 and Cy5 from each sample.

Competitive hybridizations were performed in duplicate using Cy3- and Cy5-labeled cDNA from each preparation to eliminate background noise caused by possible differences in labeling efficiency of the Cy dyes and variable genes. In brief, 20 μg of total RNA were reverse transcribed into single stranded cDNAs. During this reaction aminoallyl-labelled dUTP (aa-dUTP) was used along with other deoxynucleotides for the synthesis of cDNAs. The aa-dUTP incorporated cDNAs were then coupled with monofunctional Cy3- and Cy5-fluorescent dyes (Amersham, Uppsala, Sweden).

The 19K array was pre-hybridized in GeneMachines hybridization chambers in a 42°C water bath for at least 1 hr. Labeled cDNAs were hybridized to slides in a Maui Hyb hybridization chamber (BioMicro Systems, Salt Lake City, UT, US) at 42°C for 14 to 16 hr. Slides were scanned using GenePix4000S (Axon Instruments, Sunnyvale, CA, USA) and the resultant composite images were analyzed by GenePix Pro 4.0 software. The gene expression analysis of four pairs of the chromosome 11 microcell hybrids and their tumour segregants (HK11.8/HK11.8-3TS, HK11.12/HK11.12-2TS, HK11.13/HK11.13-1TS, and HK11.19/HK11.19-4TS) was carried out using GIS Microarray Analysis Database (GIS mAb, http://gismadb.gis.a-star.edu.sg).

1.2 Reverse transcription-polymerase chain reaction (RT-PCR) analysis of THY1

One μg of total RNA was reverse-transcribed with 200 U of SuperScript II reverse transcriptase (Invitrogen, Carlsbad, CA, US). For PCR amplifications of specific cDNAs of *THY1* gene, a pair of sense and antisense oligonucleotides was designed with the Primer3 primer design program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and the sequences were 5'-GACCCGTGAGACAAAGAAGC-3' and 5'-GCCCTCACACTTGACCAGTT-3'. In

addition, the primer sequences were chosen from separate exons of the genes so that the RT-PCR product could readily be distinguished from any genomic DNA-induced PCR product. The primer sequences of glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) were used as previously described (6). PCR was performed in a GeneAmp PCR System 9700 programmable thermal controller (Applied Biosystems, Foster City, CA, US).

1.3 Identification of the putative THY1 promoter

The putative *THY1* promoter was identified based on the study of Qiu et al.[126]. To define the *THY1* promoter, the upstream sequence of the *THY1* exon 1 was retrieved from UCSC Genome Browser (http://genome.ucsc.edu, extracted fragment NT_033899). A moderately GC-rich putative *THY1* promoter was identified by computer analysis (http://www.genomatix.de), which is 718 bp, located at 22,837,541-22,838,258 of fragment NT_033899. The 3' end region spanning this putative promoter fulfilled the criteria of a CpG island, CpG island starts at 402 bp and ends at 714 bp; GC content, 70%; observed/expected CpG ratio, 0.602; and a total of 22 CpG sites in a 313 bp region.

1.4 Methylation-specific PCR (MSP) analysis

The design of primer pairs for methylated and unmethylated DNA was based on MethPrimer (www.urogene.org/methprimer) primer design program using guidelines for MSP primer selection (8). The sequences were 5'-TATTTTTATATTAATGCGGGATCGT-3' and 5'-CGATTACTACACCCAACTCGAA-3' for methylation primers and 5'-TTATTTTTATATTAATGTGGGATTGT-3' and 5'-TCCAATTACTACACCCAACTCAAA -3' for unmethylated primers. In brief, 5 □g genomic DNA was bisulfite-modified using the

CpGenome^{TM} DNA Modification Kit (Chemicon International, Inc, Temecula, CA, US), and modified DNA was subjected to PCR. CpGenome^{TM} Universal Methylated DNA (Chemicon International, Inc, Temecula, CA, US) was used as a positive control for methylated DNA. For each PCR, controls without DNA templates were also included and showed negative results.

1.5 5-Aza-2'-deoxycytidine treatment

HONE1 cells were seeded in a T75 culture flask 24 h before treatment. Culture medium containing 5 μM 5-aza-2'-deoxycytidine (Sigma, St Louis, MO, US) was added to the cells. The medium was changed every 24 h. After incubation for 4 d, HONE1 cells were harvested for total RNA extraction.

1.6 Western blot analysis of THY1 protein

Viable NPC cells at 70-80% confluency (5×10^6) were harvested, and then lysed with 50 mM Tris (pH 7.5), 100 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, and protein inhibitor mixture (Roche, Nutley, NJ, US). Protein concentrations were measured by the Bradford assay (Bio-Rad Laboratories, Munich, Germany). Samples of 20 μg cellular protein were separated on 10% SDS-polyacrylamide gels and transferred to polyvinylidine difluoride (PVDF) membranes (pore size: 0.45 μm; Millipore, Billerica, MA, US). The membranes were blocked with 5% skim milk and primary antibody incubation was performed with 1:2500 H-110 (Santa Cruz Biotechnology, Santa Cruz, CA, US) for the THY1 protein, and 1:5000 Ab-1 (Calbiochem, Darmstadt, Germany) for α-tubulin. The signals were visualized by enhanced chemiluminescence method according to the manufacturer's instructions (Amersham, Uppsala, Sweden).

1.7 Analysis of Patient specimens

Eighty NPC specimens and 20 specimens of nasopharyngeal mucosa from non-NPC diseases were collected in the Cancer Institute of Sun Yat-Sen University, between 1997 and 2000. Ages of the 80 carcinoma patients ranged from 14 to 72 years (mean, 47 years old). The male and female ratio is 4.3:1. The tumour specimens encompassed 50 primary NPC cases without metastasis and 30 lymph node metastatic NPC. All NPC cases selected in this study were poorly differentiated squamous cell carcinomas.

1.8 NPC tissue microarray (NPC-TMA)

The TMA was constructed according to the method described previously[127]. For the construction of NPC TMA, one sample was selected from each case of the 80 NPC's (50 primary carcinomas and 30 metastatic NPC) and 20 cases of normal nasopharyngeal mucosa. Multiple sections (5 μm thick) were cut from the TMA block and mounted on microscope slides.

1.9 Immunohistochemical (IHC) staining

Immunohistochemistry studies were performed using the standard streptavidin-biotin-peroxidase complex method as described[127]. In brief, the TMA slides were incubated with rabbit anti-THY1 polyclonal antibody (H-110, Santa Cruz Biotechnology, Santa Cruz, CA, US, 1:500 dilution) overnight. The slides were then incubated with a biotinylated goat anti-rabbit serum for 30 minutes and subsequently reacted with a streptavidin-peroxidase conjugate and 3'-3' diaminobenzidine. The nucleus was counterstained using Meyer's hematoxylin. For the negative control, the primary antibody was replaced with blocking serum.

Positive expression of THY1 was primarily a cytoplasmic pattern (Fig. 3). Both staining intensity and positive areas were recorded. A staining index (values 0 to 9) obtained as the intensity of THY1-positive staining (negative = 0, weak = 1, moderate = 2, or strong = 3 scores) and the proportion of immunopositive cells of interest (<10% = 1, 10 to 50% = 2, > 50% = 3 scores), were calculated.

1.10 Gene transfection and colony formation assay

The full-length wild type *THY1* cDNA flanked with BamHI restriction sites [80] was ligated to the pCR3.1 neomycin-resistance tagged expression plasmid. The construct containing the *THY1* gene, pCR3.1-*THY1*, was confirmed by sequencing on an ABI3100 Sequencer (Applied Biosystems, Foster City, CA, US). PCR3.1-THY1 recombinant and control pCR3.1 (Invitrogen, Carlsbad, CA, US) vector-alone cell lines were transfected into HONE1 cells with Lipofectamine 2000 Reagent (Invitrogen, Carlsbad, CA, US). A total of 4×10^5 cells were seeded in 6-well plates and were transfected with 1 μg of plasmid DNA for 5 h. The cells were subsequently split into 100 mm dishes. After 14 days of selection in DMEM/10%FCS containing 400 μg/ml neomycin, colonies were fixed and stained with Giemsa to assess the transfection efficiency.

Construction of a pETE-Bsd responsive vector and a HONE1 cell line, HONE1-2, producing the tetracycline trans-activator tTA as described in Protopopov *et al.*[68]. The full-length *THY1* cDNA was inserted into BamHI digested pETE-Bsd plasmid containing a selectable blasticidin-resistance gene and the recombinant pETE-Bsd-*THY1* was verified by sequencing on ABI3100 Sequencer (Applied Biosystems, Foster City, CA, US). pETE-Bsd-*THY1* and control

pETE-Bsd vector were transfected into the HONE1-2 cells as described above. After 14 days of selection in DMEM/10%FCS containing 5 μg/ml blasticidin and 400 μg/ml neomycin with or without 0.5 μg/ml doxycycline, colonies were fixed and stained.

1.11 Statistical analysis for THY-1 expression

The X^2 and Fisher's Exact test were used for analysis of significant differences in *THY1* expression level detected by TMA between primary and metastatic NPC, and in colony numbers between control and *THY1*-transfected cells. Differences were considered statistically significant for p values < 0.05.

1.12 Quantitative PCR

Primer Express software supplied by Applied Biosystems (Darmstadt, Germany) and primerbank were used to generate the oligonucleotide sequences [128]. Real-time PCR was carried out using the LightCycler® FastStart DNA MasterPLUS SYBR Green I kit (Roche Applied Science, Germany) and Light Cycler machine according to the manufacturer's protocol. The annealing temperature was fixed at 60ºC. Melting curve analysis was carried out to determine the specificity of the primers. PFAFFL's relative quantification method was used to determine the fold changes between the hybrid's and segregants [129].

1.13 FISH

Metaphase spreads were prepared by following the published procedures. DNA probes were labeled with Texas Red (TxR) and fluorescein isothiocyanate (FITC) (Perkin Elmer) using the Bioprime labeling kit (Invitrogen) and hybridized to metaphase spreads from Parental (Hone1), Hybrids (11.8 and 11.19) and Segregants (11.8_ts and 11.19_ts) as previously

described. The slides were counter stained using 4', 6-diamidino-2-phenylindole dihydrochloride in order to visualize the chromosomes. Fluorescent hybridizations signals and DAPI staining patterns were captured using an IMAC-CCD=S36 camera attached to a Nikon 80I microscope. The images were processed using ISIS (Insitu Imaging System) image processing software (metasystems). At least 30 metaphases were examined per probe.

1.14 SKY

The SKY-Paint DNA kit from Applied Spectral Imaging (Mannheim, Germany) was hybridized to metaphase spreads according to the manufacturer's protocol. Chromosomes were counterstained with 4′, 6-diamidino-2-phenylindole (DAPI; Sigma, Taufkirch, Germany). The multicolor hybridizations were visualized with the Spectra-Cube SD300; spectral analysis and classification was done with the SKY-View 1.6 software (Applies Spectral Imaging). A minimum of 15 metaphases were examined in each case.

1.15 Modified Fisher's Exact Score

DAVID annotation system that was used for the functional annotation analysis adopted a modified fisher's exact score that was more stringent than the conventional Fisher's exact score. The difference between the two scores are depicted in the 2x2 contingency table below.

Conventional Fisher's exact Score

	Selected gene list	Background gene list (All the genes in the genome)
In Pathway	A	C
Not In Pathway	B	D

The modified Fisher's exact (Ease Score).

	Selected gene list	Background gene list (All the genes in the genome)
In Pathway	A-1	C
Not In Pathway	B	D

1.16 In-silico binding site analysis

Batch Extraction and Analysis of *cis*-Regulatory Regions (BEARR) program was used to extract and analyze the promoter region of the genes. 3kb base pairs upstream of the transcription start site were extracted and analyzed using the PWM module. The PWM for the transcription factor was extracted from the transfact database. Log-likelihood approach was adapted in the

PWM analysis. The Log-likelihood score measures the ratio of the probability that a subsequence is generated by the model (i.e. PWM) versus the probability that the subsequence is generated randomly (i.e. using the null background model). The score cutoff 0 (for log-likelihood) carries an implicit criterion that sites which are more likely to be generated by the PWM model of interest rather than by chance. Further, empirical p-value calculation enables users to easily filter the patterns found in a more principled way.

The *p*-value of each pattern found is based on the null hypothesis that the pattern (with the associated PWM score) could be found by chance alone. The *p*-value is calculated by generating 100,000 random sequences of the same length of the pattern and reporting the fraction of times the random sequence is scored better or at least as good as the observed one.

1.17 Western blot and immunofluorescence analysis

The total extracts were harvested from the subconfluent cells as described [130]. Twenty micrograms of protein was boiled in 5X sample loading buffer buffer for 5 min and separated using sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Proteins were then electroblotted onto polyvinyl difluoride membranes (Bio-Rad,), and nonspecific binding sites were blocked for 1 h at room temperature by 5% wt/vol) fat-free milk before an overnight incubation at 4°C with specific antibodies: mouse anti-human SREBP-1 (1:1,000), mouse anti-human SREBP-2 (1:1,000), and mouse anti-human SCAP (1:1000). Anti-human beta-actin antibody (1:1000; SantaCruz) was used as a loading control. Primary antibodies were detected with horseradish peroxidase-conjugated goat anti-mouse IgG (1:5000; Santa Cruz). Blots were then treated with Enhanced Chemiluminescence detection kit (Amersham,) for detecting the

protein. Note: for SREBP protein detection, cells were treated with 25 mg/ml ALLN calpain I inhibitor; Calbiochem, Merck) for the last 3 h before harvesting, in order to stabilize the short-lived nuclear forms, as described previously [131]. NPC cells were seeded in Lab-tek 8 well chamber slides. Cells were fixed in 4% Para formaldehyde for 15 min at room temperature, washed three times, blocked and permeabilized with 5% FBS in 1XPBS/0.3% triton X-100 for 60 min. Cells were then incubated with primary mouse anti-human SREBP-1 monoclonal antibody (1:250) or mouse anti-human SREBP-2 monoclonal antibody (1:250), in PBS/triton x -100 at 4C overnight. After washing, cells were incubated with FITC-conjugated anti-mouse antibody (1:750) in PBS/triton for 1-2 hours and washed three times with PBS. Immunofluorescence was analyzed using a fluorescence microscope (Nikon,).

1.18 Construction of vectors for luciferase assay

All the constructs for the luciferase assay were made using the pGL3-Basic vector from Promega. The regulatory regions of the genes were determined using the BEARR 1.0 tool. Primers were designed to extract 3000bp upstream and 500bp downstream of TSS. The primers were designed with the KpnI and SacI restriction enzyme sites to be sub-cloned into the KpnI and SacI sites in the vector (Fig 1).

1.19 Disruption of SRE binding sites in the regulatory region

PWM module of the BEARR 1.0 tool was used to determine the SRE binding sites in the genes of interest. The PWM matrix for the SRE binding sites were derived form the transfac database for the transcription factor binding sites. Overlap PCR based approach was used for the disruption of the SRE binding sites in the regulatory region of the genes. Two internal primers

running opposite to each other were designed for the immediate flanking region of the SRE binding sites. These primers were tagged with a random sequence that will replace the SRE binding sites in the promoter region. The tagging random sequences in the two primers are designed in such a way that they are complimentary to each other. Separate PCR reactions were carried out to amplify the two separate parts of the regulatory region. The products from these PCR reactions were pooled together and PCR amplified using two external primers that were designed with the KpnI and SacI sites. Thus the SRE binding sites were disrupted in the regulatory regions. These fragments were later sub-cloned in to the pGL3-Basic vector using the KpnI and SacI sites (Fig 2).

1.20 Transfection of luciferase constructs in to the NPC cells

One day prior to the transfection the NPC cells were plated at a concentration of 1-2 x 10^4 in 100 µl DEPM/10%FCS medium. The cells were allowed to reach 90-95% confluency. The DNA (6 ng of renilla luciferase construct and 150 ng of firefly luciferase construct) and the Lipofectamine 2000 were separately diluted using 15 µl of Opti-MEM I reducing medium with out serum and Opti-MEMI medium respectively. They were incubated separately at room temperature for 5 minutes before combining them together. The mixture was allowed to remain at room temperature for 20 minutes after gentle mixing. 30 µl of this mixture were added in to the wells containing the cells and the medium and mixed gently by rocking the plate back and forth. The cells were incubated for 36 hours before carrying out the luciferase assay.

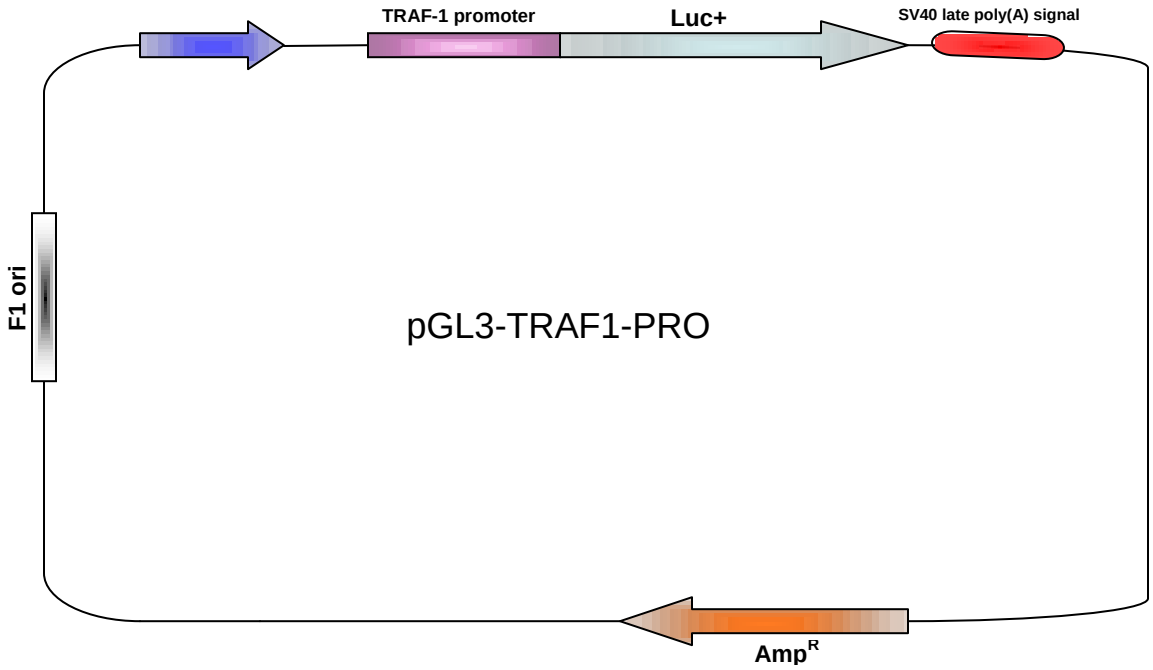


Figure A1: Construction of vectors for luciferase assay.

The Diagram above shows the vector construct using TRAF-1 promoter sequence, Similar constructs using TNFAIP3 (A20) and IKB-α were made for luciferase assay. 3000bp upstream of the TSS and 500bp downstream of TSS were subcloned with the aid of PCR. Restriction enzyme sites KpnI and SacI were introduced in to the primers used for sub cloning the regulatory sequences. These sites were later used to subclone them in to the pGL3-Basic vector (Promega). Similar constructs were made for all the three genes after disrupting the SRE sites in their promoter region.

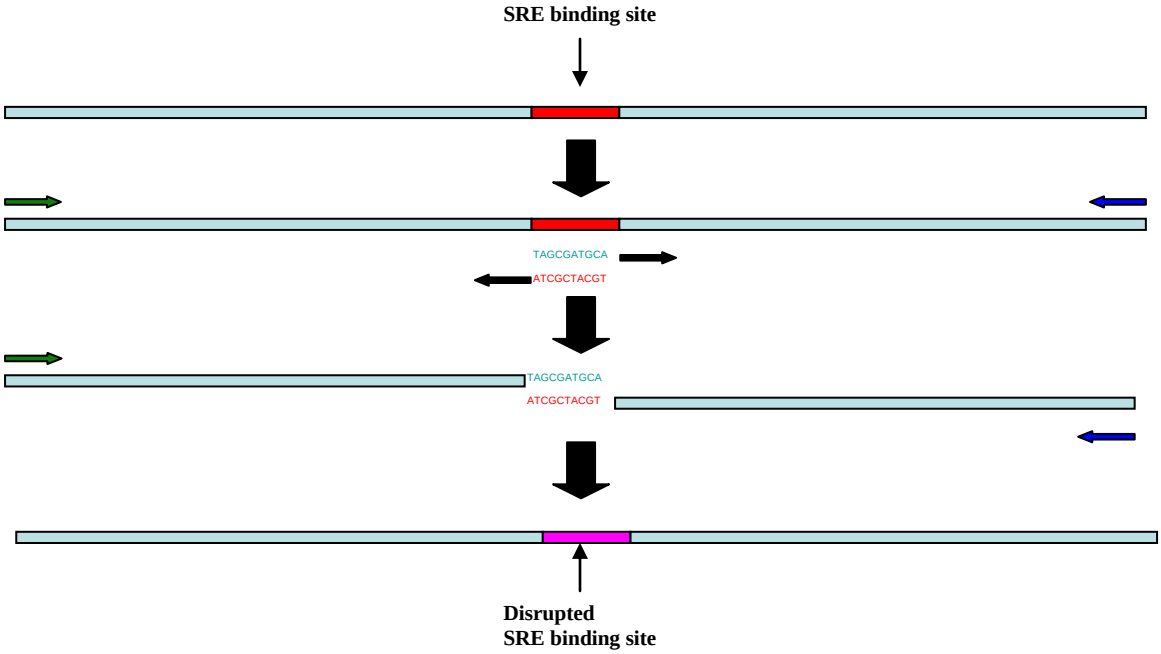


Figure A2: Disruption of SRE sites.

The SRE sites in the TRAF-1, A20 and IKB-α were disrupted with the aid of overlapping PCR. Internal primers were selected using the sequences flagging the SRE binding sites. These primers were tagged with a random sequence that differs entirely from the SRE binding site signature. These random sequences in the two internal primers are designed in such a way that they are complimentary. The two PCR reaction products containing the first half and second half are pooled and primers spanning the entire full length promoter region was used to amplify the full length sequence with the disrupted SRE binding site.

1.21 Luciferase Assay

The growth media is removed from and the cells were rinsed with 1X PBS. 20 µl 1XPLB (Passive Lysis Buffer) was dispensed in to each well containing the cells. The plates were rocked for 15 minutes at room temperature. 100 µl LARII (Luciferase assay Reagent) was dispensed in to each well and firefly luciferase activity was first measured using Centro LB960 96-well luminometer (Berthold Technologies). The firefly luciferase activity was then quenched using 100 µl of Stop and Glo reagent followed by the measurement of renilla luciferase activity.

1.22 Chromatin immunoprecipitation assays

NPC cells were cultured for 48 h till 80% confluency and treated with 1% formaldehyde to cross-link the transcription machinery and the chromatin. Immunoprecipitations were carried out overnight with SREBP or GST antibodies (Santa-Cruz Biotechnology) and protein A-sepharose beads (Zymed). Washing and extraction protocols were adapted from methods described previously [62] and PCR reactions were carried out in a LightCycler (Roche Diagnostics) real-time system. Thirty six cycles of PCR were carried out on precipitated DNA and control DNA using the respective primer sets (Annexe II). PFAFFL's relative quantification method was used to determine the fold changes between the anti-SREBP and anti-GST control. [129].

APPENDIX II

Quantitative PCR Primers

1) Insig-1_F	5’ GCCTACTGTACCCCTGTATCG
2) Insig-1_R	5’ TGGTTAATGCCAACAAAAACTGC
3) SAA1_F	5’ TCAGGTGAGGAGCACACCAA
4) SAA1-R	3’ CCAGGACCAAGGAGCAGAAA
5) SAA2_F	5’ TCAGGTGAGGAGCACACCAA
6) SAA2_R	5’ CCCGTGAGAAGCTTCATGGT
7) ABL2_F	5’ GCAAGAGGCGAATCTGGTG
8) ABL2_R	5’ CGTGGTGTAAAGAAGCCTGTG
9) ALG1_F	5’ ACAACTGACTCTTGATGGACACA
10) ALG1_R	5’ AAGCAGGGGGTAGTCCTCG
11) MRPL28_F	5’ CCTGTGCTCCAAGTTTGGGAT
12) MRPL28_R	5’ GATCAGCTCCGCCACATAGA
13) MSLN_F	5’ ACCCACCTAACATTTCCAGCC
14) MSLN_R	5’ AGGAATAGCAGCAGGTCCAAT
15) NUBP2	5’ GTGGAGAATATGAGCGGCTTC
16) NUBP2	5’ CGTCGCGTCCAGAATCTTC
17) RHOT2_F	5’ TGCCTCTTTGTCTCCTCCAAG
18) RHOT2_R	5’ CAGGACCCTGTAGAGTGAGAAG
19) TFAP4_F	5’ GCAGACAGCCGAGTACATCTT
20) TFAP4_R	5’ CCTATGCCTTCGTCCTTGTCC
21) TMEM8_F	5’ GCCTCACAAACTACCCAGTCA
22) TMEM8_R	5’ CTCGTTCCGCATCTCTGTCTT
23) ZNF434_F	5’ GAACCCACGGAGGTAGAAGAT
24) ZNF434_R	5’ GCCAATAAACTCCTCTGGACTTT
25) C1orf9_F	5’ TGCCGAATCATTGGGAAAATCA
26) C1orf9_R	5’ CCACAACTGCATTGGAAGACTC
27) HSD17B7_F	5’ AGCCTGAATCTCTCAATCCTCT

30) HSD17B7_R 5’ GCAGTGTCTTCATCTAGGTCCA

31) MRPS14_F 5’ GAATACGCAGATGAGAGGCTAC

32) MRPs14_R 5’ CGTCATAACACACCGATTTCTGA

33) RALGPS2_F 5’ GAGGCCAAGCTGAAAGTTCTAC

34) RALGPS2_R 5’ GATAGCCATTTCGTGCCACTC

Luciferase assay construct primers

35) Traf1_F 5’ GATCGCGGTACC GAGGTTGGCTGCCCTACAC

36) Traf1_R 5’ GGTCACGAGCTC AGCACCAGACCACACAGTGA

37) TNFAIP3_F 5’ GATCGCGGTACC CGTGGTGCTCTCTTTCATCA

38) TNFAIP3_R 5’ GGTCACGAGCTC AGCGCTTCTGCAAGGTCTAC

39) IKB-α_F 5’ GATCGCGGTACC GACAGGGTCTCAGTCTGTTGC

40) IKB-α_R 5’ GGTCACGAGCTC ACTTACGAGTCCCCGTCCTC

41) Traf1_SRE_MF 5’ CAAGTACAGTACGGCTCTCAGCCCTTCCGGA

42) Traf1_SRE_MR 5’ TACTGTACTTGGGCAGGAGGGCTGGTTGGAT

43) IKBA1_SRE_MF 5’ CCGGCAGACTACAACAGGTTTTCTCCCCATC

44) IKBA1_SRE_MR 5’ TAGTCTGCCGGTCAAAACTGTTTCCCAGAGT

45) IKBA2_SRE_MF 5’ AGCAGACGTGGACAATGAAAACTTCCCAGGG

46) IKBA2_SRE_MR 5’ CCACGTCTGCTATTATTACCCTGGAGGGTCT

CHIP assay RT-PCR Primers

47) TRAF1_CF 5’ GATGGAGGCCCAGTGTAGAA

48) TRAF1_CR 5’ CCCCAGTTTCTATCCAACCA

49) TNFAIP3_CF 5’ ATTTCCACGGGACTTTCCA

50) TNFAIP3_CR 5’ CGAAATGCCCAGGTGACT

51) NFKBIA1_CF 5’ GTCTTCTGCCCAGTCTGCTC

52) NFKBIA1_CR 5’ CTGCGACTCTGGGAAACAGT

53) NFKBIA2_CF 5’ CCAATCCTTCCCCTCTCTTC

54) NFKBIA2_CR 5’ TCCCTGGGAAGTTTTCATTG

55) DHCR7_CF 5’ GCCGTCAATCTCGAGTCC

56) DHCR7_CR 5’ CTAGCCAGGGGTCGGAGT

BIBLIOGRAPHY

1. Hanahan, D.a.W., R.A., *The hallmarks of cancer.* Cell, 2000. **100**(57-70).

2. Slamon, D.J., Clark, G.M., Wong, S.G., Levin, W.J., Ullrich, A. and W.L. McGuire, *Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene.* Science, 1987. **236**: p. 177-182.

3. Yarden, Y.a.U., A., *Growth factor receptor tyrosine kinases.* Annu Rev Biochem, 1988. **57**: p. 443-478.

4. Fedi, P., Tronick, S.R., and Aaronson, S.A., *Growth factors in Cancer Medicine*, ed. R.C.B. J.F. Holland, D.L. Morton, E. Frei, D.W. Kufe, and R.R. Weichselbaum,. 1997, Baltimore, MD: Williams and Wilkins. 41-64.

5. Medema, R.H.a.B., J.L., *The role of p21-ras in receptor tyrosine kinase signaling.* Crit. Rev. Oncog., 1993. **4**: p. 615-661.

6. Kinzler, K.W., and Vogelstein, B., *Cancer-susceptibility genes. Gate-keepers and caretakers.* Nature, 1997. **386**: p. 761-763.

7. Weinberg, R.A., *The retinoblastoma protein and cell cycle control.* Cell, 1995. **81**: p. 323-330.

8. Moses, H.L., Yang, E.Y. and Pietenpol, J.A., *TGF-b stimulation and inhibition of cell proliferation: a new mechanistic insights.* Cell, 1990. **63**: p. 245-247.

9. Kinzler, K.W., and Vogelstein, B., *The genetic basis of human cancer.* edn 2, Edited by Kinzler KW, Vogelstein B, Totonto: Mc Graw-Hill, 2002. **3**: p. 821.

10. Albertson, D.G., Collins, C., McCormick, F., and Gray, J. W, *Chromosome aberrations in solid tumors.* Nat Genet, 2003. **34**: p. 369-376.

11. Lerebours, F., Bertheau, P., Bieche, I., et al., *Two prognostic groups of inflammatory breast cancer have distinct genotypes.* Clin Cancer Res, 2003. **9**: p. 4184-4189.

12. Rennstam, K., Ahlstedt, S. M., Baldetorp, B., et al., *Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization.* Cancer Res, 2003. **63**: p. 8861-8868.

13. Chandrasekharappa, S.C., Guru, S. C., Manickam, P., Olufemi, S. E., Collins, F. S., Emmert-Buck, M. R., Debelenko, L. V., Zhuang, Z., Lubensky, I. A., Liotta, L. A., Crabtree, J. S., Wang, Y., Roe, B. A., Weisemann, J., Bogusk,i M. S., Agarwal, S. K., Kester, M. B., Kim, Y. S., Heppner, C., Dong, Q., Spiegel, A. M., Burns, A. L. and Marx, S. J., *Positional cloning of the gene for multiple endocrine neoplasia-type 1.* Science, 1997. **276**(5311): p. 404-407.

14. De Angelis, P.M., Stokke, T., Beigi, M., Mjaland, O. and Clausen, O.P., *Prognostic significance of recurrent chromosomal aberrations detected by comparative genomic hybridization in sporadic colorectal cancer.* Int J Colorectal Dis., 2001. **16**(1): p. 38-45.

15. Mitelman, F., *Recurrent chromosome aberrations in cancer.* Mutat rRes., 2000. **462**: p. 247-253.

16. Jones, P.A.a.B., S.B., *The fundamental role of epigenetic events in cancer.* Nat Rev Genet., 2002. **3**(6): p. 415-428.

17. Tsou, J.A., Hagen, J.A., Carpenter, C.L. and Laird-Offringa, I.A., *DNA methylation analysis: a powerful new tool for lung cancer diagnosis.* oncogene, 2002. **21**(35): p. 5450-5461.

18. Costello, J.F.a.P., C., *Methylation matters.* J Med Genet., 2001. **38**: p. 285-303.

19. Esteller, M., *CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future.* oncogene, 2002. **21**: p. 5427-5440.

20. Momparler, R.L., *Cancer epigenetics.* oncogene, 2003. **212**: p. 6479-6483.

21. Kitano, H., *Cancer robustness: Tumour tactics.* Nature, 2003. **426**: p. 6963.

22. Bignold, L.P., Coghlan, B.L. and Jersmann, H.P., *Hansemann, Boveri, chromosomes and the gametogenesis-related theories of tumours.* Cell Biol Int., 2006. **30**(7): p. 640-644.

23. Rowley, J.D., *A new consistent chromosomal abnormality in chronic myelogenous leukemiaidentified by quinacrine fluorescence and giemsa staining.* Nature, 1973. **243**: p. 290-293.

24. Waters, J.J., Barlow, A. L., and Gould, C. P., *Demystified ... FISH.* 1998. **51**: p. 62-70.

25. Ried, T., Schrock, E., Ning, Y., and Wienberg, J., *Chromosome painting: a useful art.* Human molecular genetics, 1998. **7**(10): p. 1619-1626.

26. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., et al., *Matirx-based comparative genomic hybridization: biochips to screen for genomic imbalances.* Genes Chromosomes and Cancer, 1997. **20**: p. 399-407.

27. Pinkel, D., Segraves, R., Sudar, D., et al., *High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.* Nature genetics, 1998. **20**: p. 207-211.

28. Aguirre, A.J., Brennan, C., Bailey, G., et al., *High resolution charaterization of thepancreatic adenocarcinoma genome.* Proc Natl Acad Sci USA, 2004. **101**(24): p. 9067-9072.

29. Lucito, R., Healy, J., Alexander, J., et al, *Representational oligonucleotide microarray analysis: a high resolution method to detect genome copy number variation.* Genome research, 2003. **13**: p. 2291-2305.

30. Bignell, G.R., Hunag, J., Greshock, J., et al, *High resolution analysis of DNA copy number using oligonucleotide microarrays.* Genome research, 2004. **14**: p. 287-295.

31. Zhao, X., Li, C., Paez, J. G., et al, *An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays.* Cancer research, 2004. **64**: p. 3060-3071.

32. Snijders, A.M., Nowak, N., Segraves, R., et al., *Assembly of microarrays for genome-wide measurement of DNA copy number.* Nature genetics, 2001. **29**: p. 263-264.

33. Fiegler, H., Carr, P., Douglas, E. J., et al., *DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones.* Genes Chromosomes and Cancer, 2003. **36**: p. 361-374.

34. Chung, Y.J., Jonkers, J., Kitson, H., et al., *A whole genome mouse BAC microarray with 1-mb resolution for analysis of DNA copy number changes by array comparative genomic hybridization.* Genome research, 2004. **14**: p. 188-196.

35. Albertson, D.J., Ylstra, B., Segraves, R., et al, *Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene.* Nature genetics, 2000. **25**: p. 144-160.

36. Zafarana, G., Grygalewicz, B., Gillis, A. J., et al., *12-p amplicon structure analysis in testicular germ cell tumors of adolescents and adults by array CGH.* oncogene, 2003. **22**: p. 7695-7701.

37. Emanuel, B.S.a.S., S.C., *From microscopes to microarrays: dissceting recurrent chromosomal rearrangements.* Nat Rev Genet., 2007. **8**(11): p. 869-883.

38. Holst, F., Stahl, P.R., Ruiz, C., Hellwinkel, O., Jehan, Z., Wendland, M., Lebeau, A., Terracciano, L., Al-Kuraya, K., Janicke, F., Sauter, G. and Simon, R., *Estrogen recceptor alpha (ESR1) gene amplification is frequent in breast cancer.* Nat Genet, 2007. **39**(5): p. 655-660.

39. Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., Chen, F., Feiler, H., Tokuyasu, T., Kingsley, C., Dairkee, S., Meng, Z., Chew, K., Pinkel, D., Jain, A., Ljung, B.M., Esserman, L.,

Albertson, D.G., Waldman, F.M. and Gray, J.W., *Genomic and transcriptional aberrations linked to breast cancer pathophysiologies.* Cancer Cell, 2006. **10**(6): p. 529-541.

40. Adler, A.S., Lin, M., Horlings, H., Nuyten, D.S., van de Vijver, M.J. and Chang, H.Y., *Genetic regulators of large-scale transcriptional signatures in cancer.* Nat Genet, 2006. **38**(4): p. 421-430.

41. Chin, S.F., Teschendorff, A.E., Marioni, J.C., Wang, Y., Barbosa-morais, N.L., Thorne, N.P., Costa, J.L., Pinder, S.E., Van de Wiel, M.A., Green, A.R., Ellis, I.O., Porter, P.L., Tavare, S., Brenton, J.D., Ylstra, B. and Caldas, C., *High-resolution cCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer.* Genome Biol., 2007. **8**(10): p. R215.

42. Wei, W.I.a.S., J.S., *Nasopharyngeal carcinoma.* Lancet, 2005. **365(9476):2041-54**(9476): p. 2041-2054.

43. Lu, C.C., Chen, J. C., Jin, Y. T., Yang, H. B., Chan, S. H., and Tsai, S. T., *Genetic susceptibility to nasopharyngeal carcinoma within the HLA-A locus in Taiwanese.* Cancer genetics, 2002. **103**(6): p. 145-151.

44. Armstrong, R.W., Armstrong, M. J., Yu, M. C., and Henderson, B. E., *Salted fish and inhalants as risk factors for nasopharyngeal carcinoma in Malaysian Chinese.* Cancer research, 1983. **43**: p. 2967-2970.

45. Lung, M.L., Sham, J., Lam, W. P. and Choy, D., *Analysis of localized tumors provides further evidence for the direct association of Epstein-Barr virus with nasopharyngeal carcinoma.* Cancer, 1993. **71**: p. 1190-1192.

46. Cheng, Y., Poulos, N. E., Lung, M. L., Hampton, G., Ou, B., Lerman, M. I., and Stanbridge, E. J., *Functional evidence for a nasopharyngeal carcinoma tumor suppressor gene that maps at chromosome 3p21.3.* Proc Natl Acad Sci USA, 1998. **95**(6): p. 3042-3047.

47. Huang, D.P., Lo, K. W., van Hasselt, C. A., Woo, J. K., Choi, P. H., Leung, S. F., Cheung, S. T., Cairns, P., Sidransky, D. and Lee, J. C., *A region of homozygous deletion on chromosome 9p21-22 in primary nasopharyngeal carcinoma.* Cancer research, 1994. **54**(15): p. 4003-4006.

48. Hui, A.B., Lo, K. W., Leung, S. F., Choi, P. H., Fong, Y., Lee, J. C. and Huang, D. P., *Loss of heterozygosity on the long arm of chromosome 11 in nasopharyngeal carcinoma.* Cancer research, 1996. **56**: p. 3225-3229.

49. Mutirangura, A., Pornthanakasem, W., Sriuranpong, V., Supiyaphun, P. and Voravud, N., *Loss of heterozygosity on chromosome 14 in nasopharyngeal carcinoma.* Int J Cancer, 1998. **78**(2): p. 153-156.

50. Larsson, C., Skogseid, B., Oberg, K., Nakamura, Y. and Nordenskjold, M., *Multiple endocrine neoplasia type 1 gene maps to chromosome 11 and is lost in insulinoma.* Nature, 1988. **332**(6159): p. 85-87.

51. Cheng, Y., Chakrabarti, R., Garcia-Barcelo, M., Ha, T. J., Srivatsan, E. S., Stanbridge, E. J., and Lung, M. L., *Mapping of nasopharyngeal carcinoma tumor-suppressive activity to a 1.8-megabase region of chromosome band 11q13.* Genes Chromosomes and Cancer, 2002. **34**(1): p. 97-103.

52. Robertson, G.P., Goldberg, E. K., Lugo, T. G. and Fountain, J. W., *Functional localization of a melanoma tumor suppressor gene to a small (< or = 2 Mb) region on 11q23.* oncogene, 1999. **18**: p. 3173-3180.

53. Carter, S.L., Negrini, M., Baffa, R., Gillum, D. R., Rosenberg, A. L., Schwartz, G. F. and Croce, C. M., *Loss of heterozygosity at 11q22-q23 in breast cancer.* Cancer research, 1994. **54**: p. 6270-6274.

54. Foulkes, W.D., Campbell, I. G., Stamp, G. W., and Trowsdale, J., *Loss of heterozygosity and amplification on chromosome 11q in human ovarian cancer.* Br J Cancer, 1993. **67**: p. 268-273.

55. Rasio, D., Negrini, M., Manenti, G., Dragani, T. A. and Croce, C. M., *Loss of heterozygosity at chromosome 11q in lung adenocarcinoma: identification of three independent regions.* Cancer research, 1995. **55**: p. 3988-3991.

56. Bethwaite, P.B., Koreth, J., Herrington, C. S. and McGee, J.O., *Loss of heterozygosity occurs at the D11S29 locus on chromosome 11q23 in invasive cervical carcinoma.* Br J Cancer, 1995. **71**: p. 814-818.

57. Shaw, M.E., and Knowles, M. A., *Deletion mapping of chromosome 11 in carcinoma of the bladder.* Genes Chromosomes and Cancer, 1995. **13**: p. 1-8.

58. Keldysh, P.L., Dragani, T. A., Fleischman, E. W., Konstantinova, L. N., Perevoschikov, A. G., Pierotti, M. A., Della Porta, G. and Kopnin, B. P., *11q deletions in human colorectal carcinomas: cytogenetics and restriction fragment length polymorphism analysis.* Genes Chromosomes and Cancer, 1993. **6**: p. 45-50.

59. Dahiya, R., Lee, C., McCarville, J., Hu, W., Kaur, G., and Deng, G., *High frequency of genetic instability of microsatellites in human prostatic adenocarcinoma.* Int J Cancer, 1997. **72**: p. 762-767.

60. Kuramochi, M., Fukuhara, H., Nobukuni, T., Kanbe, T., Maruyama, T., Ghosh, H. P., Pletcher, M., Isomura, M., Onizuka, M., Kitamura, T., Sekiya, T., Reeves, R. H. and Murakami, Y., *TSLC1 is a tumor-suppressor gene in human non-small-cell lung cancer.* Nature genetics, 2001. **27**: p. 427-430.

61. Lung, H.L., Cheng, Y., Kumaran, M. K., Liu, E. T., Murakami, Y., Chan, C. Y., Yau, W. L., Ko, J. M., Stanbridge, E. J. and Lung, M. L., *Fine mapping of the 11q22–23 tumor suppressive region and involvement of TSLC1 in nasopharyngeal carcinoma.* Int J Cancer, 2004. **112**: p. 628-635.

62. Mutirangura, A., Tanunyutthawongese, C., Pornthanakasem, W., Kerekhanjanarong, V., Sriuranpong, V., Yenrudi, S., Supiyaphun, P., and Voravud, N., *Genomic alterations in nasopharyngeal carcinoma: loss of heterozygosity and Epstein-Barr virus infection.* Br J Cancer, 1997. **76**: p. 770-776.

63. Komatsu, K., Kodama, S., Okumura, M., Koi, M. and Oshimura, M., *Restoration of radiation in ataxia-telangiectasia cells by the introduction of normal human chromosome 11.* Mutat rRes., 1990. **235**(2): p. 59-63.

64. Whitney, M., Thayer, M., Reifsteck, C., Olson, S., Smith, L., Jakobs, P.M., Leach, R., Naylor, S., Joenje, H. and Grompe, M., *Microcell mediated chromosome transfer maps the Fanconi anaemia group D gene to chromosome 3p.* Nature Genetics, 1995. **11**: p. 341-343.

65. Doherty, A.M., and Fisher, E. M., *Microcell-mediated chromosome transfer (MMCT): small cells with huge potential.* Mammalian Genome, 2003. **14**: p. 583-592.

66. Meaburn, K.J., Parris, C. N. and Bridger, J. M., *The manipulation of chromosomes by mankind: the uses of microcell-mediated chromosome transfer.* Chromosoma, 2005(114): p. 263-274.

67. Cheng, Y., Stanbridge, E. J., Kong, H., Bengtsson, U., Lerman, M. I., and Lung, M. L., *A functional investigation of tumor suppressor gene activities in a nasopharyngeal carcinoma cell line HONE1 using a monochromosome transfer approach.* Genes Chromosomes and Cancer, 2000. **28**(1): p. 82-91.

68. Protopopov, A.I., Li, J., Winberg, G., Gizatullin, R. Z., Kashuba, V. I., Klein, G., and Zabarovsky, E. R., *Human cell lines engineered for tetracycline-regulated expression of tumor suppressor candidate genes from a frequently affected chromosomal region, 3p21.* J Gene Med, 2004. **4**(4): p. 397-406.

69. Barda-Saad, M., Rozenszajn, L. A., Ashush, H., Shav-Tal, Y., Ben Nun, A., and Zipori, D., *Adhesion molecules involved in the interactions between early T cells and mesenchymal bone marrow stromal cells.* Exp Hematol, 1999. **27**: p. 834-844.

70. Diaz-Romero, J., Gaillard, J. P., Grogan, S. P., Nesic, D., Trub, T., and Mainil-Varlet, P., *Immunophenotypic analysis of human articular chondrocytes: changes in surface markers associated with cell expansion in monolayer culture.* J Cell Physiol, 2004. **202**(3): p. 731-742.

71. Cavallaro, U., and Christofori, G., *Cell adhesion in tumor invasion and metastasis: loss of the glue is not enough.* Biochem Biophys Acta, 2001. **1552**: p. 39-45.

72. Bonewald, L., Ades, E. W., Tung, E., Marchalonis, J. J., and Wang, A. C., *Biochemical characterization of THY1.* J Immunogenet, 1984. **11**: p. 283-296.

73. Crawford, J.M., and Barton, R. W., *THY1 glycoprotein: structure, distribution, and ontogeny.* Lab Invest, 1986. **54**(122-135).

74. Gunter, K.C., Malek, T. R., and Shevac, E. M., *T cell activating properties of an anti-THY1 monoclonal antibody: possible analogy to OKT3/Leu-4.* J Exp Med, 1984. **159**: p. 716-730.

75. MacDonald, H.R., Bron, C., Rousseaux, M., Horvath, C., and Cerottini, J. C., *Production and characterization of monoclonal anti-THY1 antibodies that stimulate lymphokine production by cytolytic T cell clones.* Eur J Immunol, 1985. **15**(495-501).

76. Nakashima, I., Zhang, Y. H., Rahman, S. M., Yoshida, T., Isobe, K., Ding, L. N., Iwamoto, T., Hamaguchi, M., Ikezawa, H., and Taguchi, R., *Evidence of synergy between THY1 and CD3/TCR complex in signal delivery to murine thymocytes for cell death.* J Immunol, 1991. **147**: p. 1153-1162.

77. Tentori, L., Pardoll, D. M., Zuniga, J. C., Hu-Li, J., Paul, W. E., Bluestone, J. A., and Kruisbeek, A. M., *Proliferation and production of IL-2 and B cell stimulatory factor I/IL-4 in early fetal thymocytes by activation through THY1 and CD3.* J Immunol, 1988. **140**(1089-1094).

78. Hueber, A.O., Raposo, G., Pierres, M., and He, H. T., *THY1 triggers mouse thymocyte apoptosis through a bcl-2-resistant mechanism.* J Exp Med, 1994. **179**: p. 785-796.

79. Abeysinghe, H.R., Cao, Q., Xu, J., Pollock, S., Veyberman, Y., Guckert, N. L., Keng, P., and Wang, N., *THY1 expression is associated with tumor suppression of human ovarian cancer.* Cancer Genet Cytogenet, 2003. **143**: p. 125-132.

80. Abeysinghe, H.R., Pollock, S. J., Guckert, N. L., Veyberman, Y., Keng, P., Halterman, M., Federoff, H. J., Rosenblatt, J. P., and Wang, N., *The role of the THY1 gene in human ovarian cancer suppression based on transfection studies.* Cancer Genet Cytogenet, 2004. **149**(1): p. 1-10.

81. Karnik, P., Paris, M., Williams, B.R., Casy, J., Crowe, J. and Chen, P., *Two distinct tumor suppressor loci within chromosome 11p15 implicated in breast cancer progression and metastasis.* Human molecular genetics, 1998. **7**(5): p. 895-903.

82. Li, J., Liu, Z., Wang, Y., Yu, Z., Wang, M., Zhan, Q. and Liu, Z., *Allelic imbalance of chromosome 1q in esophageal squamous cell carcinomas from China: a novel region of allelic loss and significant association with differentiation.* Cancer Lett., 2005. **220**(2): p. 221-230.

83. Naylor, T.L., Greshock, J., Wang, Y., Colligon, T., Yu, Q.C., Clemmer, V., Zaks, T.Z. and Weber, B.L., *High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization.* Breast Cancer Res., 2005. **7**(6): p. R1186-1198.

84. Dennis, G.J., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. and Lempicki, R.A., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol., 2003. **4**(5): p. P3.

85. Hosack, D.A., Dennis, G. Jr., Sherman, B.T., Lane, H.C. and Lempicki, R.A., *Identifying biological themes within lists of genes with EASE.* Genome Biol., 2003. **4**(10): p. R70.

86. Madhvi, B.U., Jens, K. H., Lisa, M. M., Edward, L. K., Carl, B. J., Michael, J. D., and Thomas Ried, *Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized cells.* Cancer Res, 2004. **64**: p. 6941-6949.

87. Kaneda, A., Kaminishi, M., Nakanishi, Y., Sugimura, T. and Ushijima, T., *Reduced expression of the insulin-induced protein 1 and p41 Arp2/3 complex genes in human gastric cancers.* Cancer genetics, 2002. **100**(1): p. 57-62.

88. Dif, N., Euthine, V., Gonnet, E., Laville, M., Vidal, H. and Lefai, E., *Insulin activates human sterol-regulatory-element-binding protein-1c (SREBP-1c) promoter through SRE motifs.* Biochem J., 2006. **400**(Pt1): p. 179-188.

89. De Smet, F., Mathys, J., Marchal, K., Thijs, G., De Moor, B. and Moreau, Y., *Adaptive qulaity-based clustering of gene expression profiles.* Bioinformatics, 2002. **18**(5): p. 735-746.

90. Sakakura, Y., Shimano, H., Sone, H., Takahashi, A., Inoue, K., Toyoshima, H., Suzuki, S. and Yamada, N., *Sterol Regulatory Element-Binding Proteins Induce an Entire Pathway of Cholesterol Synthesis.* Biochemical and Biophysical Research Communications, 2001. **286,**: p. 176–183.

91. Brown, M.S., and Goldstein, J. L., *The SREBP Pathway: Regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor.* Cell, 1997. **89**: p. 331–340.

92. Hua, X., Sakai, J., Brown, M. S., and Goldstein, J. L., *Regulated Cleavage of Sterol Regulatory Element Binding Proteins Requires Sequences on Both Sides of the Endoplasmic Reticulum Membrane.* J Biol Chem, 1996. **271,NO 17**(April 26): p. 10379–10384.

93. Sate, R., Yang, J., Wang, X., Evans, M.J., Ho, Y.K., Goldstein, J. L. and Brown, M.S., *Assignment of the membrane attachment, DNA binding and transcriptional activation domains of Sterol Regulatory Element-Binding Protein 1 (SREBP-1).* THE JOURNAL OF BIOLOGICAL CHEMISTRY, 1994. **269**: p. 17267-17273.

94. Wang, X., Zelenski, N. G., Yang, J., Sakai, J., Brown, M. S. and Goldstein, J. L., *Cleavage of sterol regulatory element binding proteins (SREBPs) by CPP32 during apoptosis.* The EMBO Journal, 1996. **15**(2): p. 1012-1020.

95. Wang, C.Y., Mayo, M.W. and ,Baldwin A.S. Jr, *TNF- and cancer therapy-induced apoptosis: potentiation by inhibition of NF-kappaB.* Science, 1996. **274**(5288): p. 784-787.

96. Logette, E., Jossic-Corcos, C. Le., Masson, D., Solier, S., Sequeira-Legrand, A., Dugail, I., Lemaire-Ewing, S., Desoche, L., Solary, E., and Corcos, L., *Caspase-2, a Novel Lipid Sensor under the Control of Sterol Regulatory Element Binding Protein 2.* Molecular and Cellular Biology, 2005. **25**(21): p. 9621–9631.

97. Wang, C.Y., Mayo, M.W., Korneluk, R.G., Goeddel, D.V. and Baldwin, A.S. Jr., *AF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation.* Science, 1998. **281**(5383): p. 1680-1683.

98. Fujioka, S., Sclabas, G.M., Schmidt, C., Frederick, W.A., Dong, Q.G., Abbruzzese, J.L., Evans, D.B., Baker, C. and Chiao, P.J., *Function of nuclear factor kappaB in pancreatic cancer metastasis.* Clin Cancer Res, 2003. **9**(1): p. 346-354.

99. Ko, H.M., Seo, K.H., Han, S.J., Ahn, K.Y., Choi, I.H., Koh, G.Y., Lee, H.K., Ra, M.S. and Im, S.Y., *Nuclear factor kappaB dependency of platelet-activating factor-induced angiogenesis.* Cancer Res, 2002. **62**(6): p. 1809-1814.

100. Ni, J., Takayama, K., Inoshima, N., Uchino, J., Harada, A., Minami, T., Harada, T., Zhou, C. and Nakanishi, Y., *Gene transfer of inhibitor kappaB in human lung cancer cell*

line NCI-H460 inhibits tumorigenesis and angiogenesis in vivo. Anticancer Res., 2005. **25**(1A): p. 69-77.

101. Erdyni, N., Tsitsikov, D.L, Ian, F.D., Tatyana, Y. S., Laurie, D., Frederick, W.A. and Raif, S.G., *TRAF1 is a negative regulator of TNF signaling: Enhanced TNF signaling in TRAF-1-deficient mice.* Immunity, 2001. **15**(4): p. 647-657.

102. Goeddel, G., Guoqing, C. and David, V., *TNF-R1 signaling: A beautiful pathway.* Science, 2002. **296**(5573): p. 1634-1635.

103. Heyninck, K.a.B., R., *A20 inhibits NF-kappaB activation by dual ubiquitin-editing functions.* Trends Biochem Sci., 2005. **30**(1): p. 1-4.

104. Baeuerle, P.A.a.B., D., *I kappa B: a specific inhibitor of the NF-kappa B transcription factor.* Science, 1988. **242**(4878): p. 540-546.

105. Zhang, B., Guan, C.C., Chen, W.T., Zhang, P., Yan, M., Shi, J.H., Qin, C.L. and Yang, Q., *A20 inhibits human salivary adenoid cystic carcinoma cells invasion via blocking nuclear factor-kappaB activation.* Chin Med J (Engl), 2007. **120**(20): p. 1830-1835.

106. Beauparlant, P., Kwan, I., Bitar, R., Chou, P., Koromilas, A.E., Sonenberg, N. and Hiscott, J., *Disruption of I kappa B alpha regulation by antisense RNA expression leads to malignant transformation.* oncogene, 1994. **9**(11): p. 3189-3197.

107. Mark, A.M., and Gregory, S. M., *Statins and cholesterol lowering after a cancer diagnosis: Why not?* Urologic Oncology, 2005. **23**: p. 49-55.

108. Coogan, P.F., Rosenberg, L. and Strom, B.L., *Statin and the risk of 10 cancers.* Epidemiology, 2007.

109. Platz, E.A., Leitzmann, M.F., Visvanathan, K., Rimm, E.B., Stampfer, M.J., Willett, W.C. and Giovannucci, E., *Statin drugs and risk of advanced prostate cancer.* J Natl Cancer Inst., 2006. **98**(24): p. 1819-1825.

110. Aerts, S., Thijs, G., Coesens, B., Staes, M., moreau, Y. and De Moor, B., *TOUCAN: deciphering the cis-regulatory logic of coregulated genes.* Nucleic Acids Research, 2003. **31**: p. 1753-1764.

111. Zhang, H., Ramanathan, Y., Soteropoulos, P., Recce, M. L., and Toias, P. P., *EZ-Retrieve: a web server for batch retrieval of coordinate-specific human DNA sequences and underscoring putative transcription factor-binding sites.* Nucleic Acids Research, 2002. **30**: p. e121.

112. Wingender, E., Dietze P, Karas H, and Knuppel R, *TRANSFAC: a database on transcription factors and their DNA binding sites.* Nucleic Acids Research, 1996. **28**: p. 316-319.

113. Bailey, T.L., and Elkan, C., *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* Proceedings of the second international conference on intelligent systems for molecular biology, 1994: p. 28-36.

114. Liu, X., Brutlag, D. L., and Liu, J. S., *Bioprospector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes.* Proceedings of the Symposium on Biocomputing, 2001: p. 127-138.

115. Liu X, B.D.L., and Liu J.S, *An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments.* Nature Biotech, 2002. **20**: p. 835-839.

116. Bussemaker, H.J., Li, H. and Siggia, E.D., *Regulatory element detection using correlation with expression.* Nat Genet, 2001. **27**: p. 167-171.

117. Pruitt, K.D.a.M., D.R., *RefSeq and Locuslink: NCBI gene-centered resources.* Nucleic Acids Research, 2001. **29**: p. 137-140.

118. Suzuki, Y., Yamashita, R., Nakai, K. and Sugano, S., *Database of human transcriptional start sites and full-length cDNAs.* Nucleic Acids Research, 2002. **30**: p. 328-331.

119. Ghosh, D., *Status of Transcription Factor Database (TFD).* Nucleic Acids Research, 1993. **21**: p. 3117-3118.

120. Kolchanov, N.A., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Stepanenko, I.L., Merkulova, T.I., Pozdnyakov, M.A., Podkolodny, N.L., Naumochkin, A.N. and Romashchenko, A.G., *Transcription Regulatory Regions Database (TRRD): its status in 2002.* Nucleic Acids Research, 2002. **30**: p. 312-317.

121. Stormo, G.D., *Consensus patterns in DNA.* Methods Enzymol, 1990. **183**: p. 211-221.

122. Clarke, P.A., te Poele, R., Wooster, R. and Workman, P., *Gene expression microarray analysis in cancer biology, pharmacology, and drug development: progress and potential.* Biochemical pharmacology, 2001. **62**(10): p. 1311-1336.

123. Masood, S., and Bui, M. M., *Prognostic and predictive value of HER2/neu oncogene in breast cancer.* Microsc Res Tech, 2002. **59**: p. 102-108.

124. Lung, H.L., Bangarusamy, D. K., Xie, D., Cheung, K. L., Cheng, Y., Kumaran, M. K., Miller, L., Liu, E. T., Guan, G., Sham, J. S., Fang, Y., Li, L., Wang, N., Protopopov, A. I., Zabarovsky, E. R., Tsao, S. W., Stanbridge, E. J., and Lung, M. L., *THY1 is a candidate tumour suppressor gene with decreased expression in metastatic nasopharyngeal carcinoma.* oncogene, 2005. **24**(43): p. 6525-6532.

125. Yao K.T., Z.H.Y., Zhu H.C., Wang F.X., Li G.Y., Wen D.S., Li Y.P., Tsai C.H., Glaser R., *Establishment and characterization of two epithelial tumor cell lines (HNE-1 and HONE-1) latently infected with Epstein-Barr virus and derived from nasopharyngeal carcinomas.* Int J Cancer, 1990. **45**(1): p. 83-89.

126. Qiu, G.H., Tan, L. K., Loh, K. S., Lim, C. Y., Srivastava, G., Tsai, S. T., Tsao, S. W., and Tao, Q., *The candidate tumor suppressor gene BLU, located at the commonly deleted region 3p21.3, is an E2F-regulated, stress-responsive gene and inactivated by both epigenetic and genetic mechanisms in nasopharyngeal carcinoma.* Oncogene. **23**: p. 4793-4806.

127. Xie, D., Sham, J. S., Zeng, W. F., Lin, H. L., Che, L. H., Wu, H. X., Wen, J. M., Fang, Y., Hu, L., and Guan, X. Y., *Heterogeneous expression and association of beta-catenin, p16 and c-myc in multistage colorectal tumorigenesis and progression detected by tissue microarray.* Int J Cancer, 2003. **107**: p. 896-902.

128. Xiaowei Wang, B.S., *A PCR primer bank for quantitative gene expression analysis.* Nucleic Acids Research, 2003. **31**(24 :e154): p. 1-8.

129. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR.* Nucleic Acids Research, 2001. **29**(9 **00**): p. 2002-2007.

130. E. Logette, C.L.J.-C., D. Masson, S. Solier, A. Sequeira-Legrand, I. Dugail, S. Lemaire-Ewing, L. Desoche, E. Solary, and L. Corcos, *Caspase-2, a Novel Lipid Sensor under the Control of Sterol Regulatory Element Binding Protein 2.* Molecular and Cellular Biology, 2005. **25**(21): p. 9621–9631.

131. Hirano, Y., M. Yoshida, M. Shimizu, and R. Sato., *Direct demonstration of rapid degradation of nuclear sterol regulatory element-binding proteins by the ubiquitin-proteasome pathway.* J. Biol. Chem., 2001. **2001**(276): p. 36431–36437.

PUBLICATIONS

1. Paul van den IJssel, Marianne Tijssen, Suet-Feung Chin, Paul Eijk, Beatriz Carvalho, Erik Hopmans, Henne Holstege, **Dhinoth Kumar Bangarusamy**, Jos Jonkers, Gerrit A. Meijer, Carlos Caldas and Bauke Ylstra. (2005) Human and mouse oligonucleotide-based array CGH ". Nucleic Acids Research 33(22):E192.

2. Lung HL, **Bangarusamy DK**, Xie D, Cheung AK, Cheng Y, Kumaran MK, Miller L, Liu ET, Guan XY, Sham JS, Fang Y, Li L, Wang N, Protopopov AI, Zabarovsky ER, Tsao SW, Stanbridge EJ, Lung ML. (2005) "THY1 is a candidate tumour suppressor gene with decreased expression in metastatic nasopharyngeal carcinoma ". Oncogene 24(43):6525-32.

3. Lin CY, Strom A, Vega VB, Kong SL, Yeo AL, Thomsen JS, Chan WC, Doray B, **Bangarusamy DK**, Ramasamy A, Vergara LA, Tang S, Chong A, Bajic VB, Miller LD, Gustafsson JA, Liu ET. (2004) "Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells". Genome Biology 5(9):R66. Epub .

4. **Bangarusamy DK** , Vega VB, Miller LD, Liu ET, Lin CY. (2004)"BEARR: Batch Extraction and Analysis of cis-Regulatory Regions". Nucleic Acids Research 32(Web Server Issue): W257-60.

5. Guy GR, Yusoff P, **Bangarusamy D**, Fong CW, Wong ES. (2002) "Dockers at the crossroads" Cell Signal 14(1):11-20.

6. Philippe Broët, Sophie Camilleri-Broët, Shenli Zhang, Marco Alifano, **Dhinoth Bangarusamy,** Maxime Battistella, Yonghui Wu, Jean-François Régnard, Elaine Lim, Patrick Tan, Lance D. Miller. Integrative analysis of genomic and transcriptional information reveals a chromosomally-driven gene signature for predicting outcome in stage IB lung carcinomas (Submitted to PNAS).

7. Arthur Kwok Leung Cheung, Hong Lok Lung, Siu Chun Hung, Evan Wai Lok Law, Yue Cheng, Wing Lung Yau1, **Dhinoth Kumar Bangarusamy**, Lance D. Miller, Edison Tak-Bun Liu, Jian-yong Shao, Chang-Wei Kou, Daniel Chua, Eugene R. Zabarovsky, Sai Wah Tsao, Eric J. Stanbridge, and Maria Li Lung. Functional analysis of a cell cycle-associated, tumor-suppressive gene, *Protein Tyrosine Phosphatase Receptor type G* (*PTPRG*), in nasopharyngeal carcinoma (Submitted to Cancer Research).

8. Mandar R. Godge, Rengasamy Ramamoorthy, **B. Dhinoth Kumar**, Ravindran, Vijay Bhaskar, Srinivasan Ramchandran, Prakash P. Kumar. Biomass and grain yield increase in rice by antisense suppression of a gene for a cytokinin binding protein (Submitted to Nature Biotech).

9. **Bangarusamy DK,** Karuthuri K, Lung ML, Miller LD, Liu ET. Carcinogenic role of cholesterol in the nasopharyngeal carcinoma cells. (Manuscript under preparation).